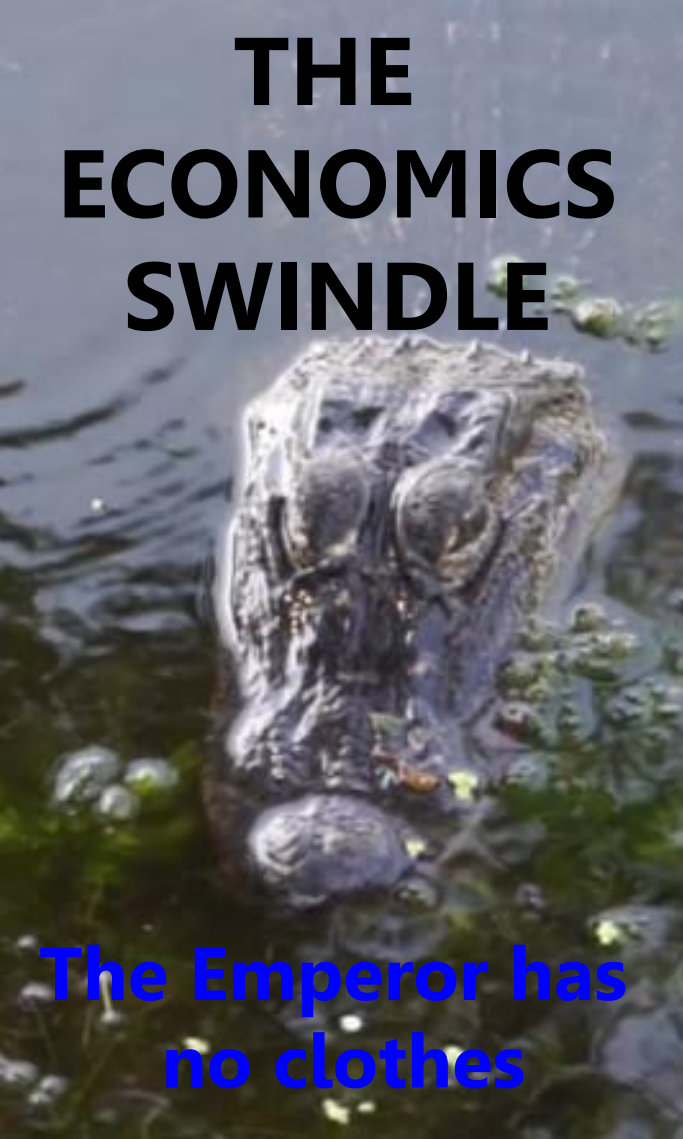
THE ECONOMICS SWINDLE



The Emperor has no clothes

Table of Contents

1. No more Mr Nice Guy. 1
2. The calculus of hedonism.. 23
3. The price of everything and the value of nothing. 57
4. Size does matter 91
5. To each according to his contribution. 125
6. The holy war over capital 146
7. There is madness in their method. 166
8. Let's do the Time Warp again. 184
9. The sum of the parts. 209
10. The price is not right 237
11. Finance and economic breakdown. 267
12. Don't shoot me, I'm only the piano. 283
13. Nothing to lose but their minds. 294
14. There are alternatives. 328

Preface

In the preface to the General Theory Keynes commented that its writing had involved a long process of escape from "habitual modes of thought and expression". He implored his audience of professional economists to likewise escape the confines of conventional economic thought, and observed that

> The ideas which are here expressed so labouriously are extremely simple and should be obvious. The difficulty lies, not in the new ideas, but in escaping from the old ones, which ramify, for those brought up as most of us have been, into every corner of our minds. (Keynes 1936)

This statement was unfortunately prophetic. Keynes's own escape was incomplete, and the residue of traditional thought the General Theory contained obscured many of its most innovative aspects. Faced with a melange of the new and unfamiliar with the old and familiar, the bulk of his audience found it easier to interpret his new ideas as no more than embellishments to the old. The Keynesian Revolution died, slowly but surely, as economists reconstructed the "habitual modes of thought and expression" around the inconvenient intrusions Keynes had made into economic dogma. Economics failed to make the escape which Keynes had implored it to do, and as time went on, 'modern' economics began to resemble more and more closely the 'old ideas' which Keynes had hoped economics would abandon.

I was initially educated in this resurgent tradition – known as the Keynesian-Neoclassical synthesis – some thirty years ago. The catalyst for my escape from this dogma was extremely simple: my first-year microeconomics lecturer pointed out a simple but glaring flaw in the application of conventional theory.

The economic theory of markets argues that combinations of any sort, whether by workers into unions or manufacturers into monopolies, reduce social welfare. The theory therefore leads to the conclusion that the world would be better off without monopolies and unions. If we were rid of both, then the economic theory of income distribution argues that, effectively, people's incomes would be determined solely by their contribution to society. The world would be both efficient and fair.

But what if you have both monopolies and unions? Will getting rid of just one make the world a better place?

The answer is categorically no. If you abolish just unions, then according to 'conservative' economic theory, workers will be exploited: they will get substantially less than their contribution to society (equally, if you abolish just monopolies, then workers will exploit companies). If you have one, then you are better off to have the other too, and a single step towards the economist's Nirvana takes you not closer to Heaven but towards Hell.[1]

I was struck by how fragile the outwardly impregnable theory of economics was. What seemed self-evident at a superficial level – that social welfare would rise if unions or monopolies were abolished – became problematic, and even contradictory, at a deeper level.

Had I come across that fragility in my honours or postgraduate education, which is when students of economics normally learn of such things, I would quite possibly have been willing to gloss over it, as most economists do. Instead, because I learnt it 'out of sequence', I was immediately suspicious of the simplistic statements of economic principle. If the pivotal concepts of competition and income distribution could be so easily overturned, what else was rotten in the House of Economics?

That scepticism initiated a gradual process of discovery, which made me realise that what I had initially thought was an education in economics was in fact little better than an indoctrination. More than a decade before I became an undergraduate, a major theoretical battle had broken out over the validity of economic theory. Yet none of this turned up in the standard undergraduate or honours curriculum – unless it was raised by some dissident instructor. There were also entire schools of thought which were antithetical to conventional economics, which again were ignored unless there was a dissident on staff.

Thirty years after starting my sceptics' intellectual tour, I am completely free of the "habitual modes of thought and expression" which so troubled Keynes. There are many non-orthodox economists like me, who are all trying to contribute to a new, deeper approach to economics.

But still the world's universities churn out economists who believe, for example, that the world would be a better place if we could just get rid of unions, or monopolies.

Worse still, over the last thirty years, politicians and bureaucrats the world over have come to regard economic theory as the sole source of wisdom about the manner in which a modern society should be governed. The world has been remade in the economist's image.

This ascendancy of economic theory has not made the world a better place. Instead, it has made an already troubled society worse: more

unequal, more unstable, and less 'efficient'.

Why has economics persisted with a theory which has been comprehensively shown to be unsound? Why, despite the destructive impact of economic policies, does economics continue to be the toolkit which politicians and bureaucrats apply to almost all social and economic issues?

The answer lies in the way economics is taught in the world's universities.

When I became an academic economist, I realised that very few of my colleagues had any knowledge of the turbulent streams in economics. Most were simply dismissive of any attempt to criticise orthodox thinking, and equally dismissive of any of their peers who showed tendencies towards unconventional thought.

This was not because these conventional economists were anti-intellectual – far from it. Even though conventional economics is flawed, it still takes intellectual muscle to master its principles – as you will soon discover. Yet still economists refused to consider any criticisms of economic theory, even when they emanated from other economists, and met rigorous intellectual standards.

Nor were they ill-intentioned – most of them sincerely believed that, if only people followed the principles of economic theory, the world would be a better place. For a group of people who espoused a philosophy of individualistic hedonism, they were remarkably altruistic in their commitment to what they saw as the common good. Yet the policies they promoted often seem to non-economists to damage the fabric of human society, rather than to enhance it.

They also rejected out of hand any suggestion that they were ideologically motivated. They were scientists, not political activists. They recommended market solutions, not because they were personally pro-capitalist, but because economic theory proved that the market was the best mechanism by which to determine economic issues. Yet virtually everything they recommended at least appeared to favour rich over poor, capitalist over worker, privileged over dispossessed.

I came to the conclusion that the reason they displayed such anti-intellectual, apparently socially destructive, and apparently ideological behaviour lay deeper than any superficial personal pathologies. Instead, the way in which they had been educated had given them the behavioural traits of zealots rather than of dispassionate intellectuals.

As anyone who has tried to banter with an advocate of some esoteric religion knows, there is no point trying to debate fundamental beliefs with a zealot. After many similar experiences with economists, I abandoned any delusion that I might be able to persuade committed economists to see reason (though there has been the odd exception to this rule). Instead, I prefer to spend my time developing an alternative approach to economics, while persuading others not to fall for the superficially persuasive but fundamentally flawed arguments of conventional theory.

Hence this book, which is aimed at a broader audience than Keynes's target of his fellow economists. Instead, my primary target market is those people who feel that they have been effectively silenced by economists. One of the many reasons why economists have succeeded in taking over social policy is that they have claimed the high intellectual ground against anyone who opposed their recommendations. The object of this book is to show that this claim is spurious.

Though I am the sole author, and thus responsible for all its errors and omissions, I cannot claim sole credit for what is good in it. In particular, I owe an enormous debt to the pioneers of critical thinking in economics.

Pre-eminent amongst these is Piero Sraffa – a name which is known to almost no non-economists, and very few economists. There are many others whose names turn up in subsequent pages – Blatt, Garengani, Goodwin, Kalecki, Kaldor, Keynes, Minsky, Veblen, to name a few. But none has had quite the impact of Sraffa.

I owe a more personal debt to those few teachers who were, as I am now, dissidents in a sea of believers. Pre-eminent here is Frank Stilwell – the first year lecturer who, many years ago, introduced me to the first of many flaws in conventional economics. I also gratefully acknowledge the influence which Ted Wheelwright's panoptic knowledge of the many currents in economic thought had upon my intellectual development. My colleagues in HETSA, the History of Economic Thought Society of Australia, have also enriched my appreciation of the many 'roads not taken' by mainstream economics.

Colleagues around the world have provided feedback on the arguments presented here. None can be held liable for what follows, but all influenced it, either directly, in debate, or by providing a forum in which heterodox views could flourish. My thanks go to Trond Andresen, George Argyrous, Tony Aspromorgous, Joanne Averill, Aldo Balardini, Bill Barnett, James Dick, Marchessa Dy, Geoff Fishburn, John Gelles, Ric Holt, Julio Huato, Alan Isaac, James Juniper, Gert Kohler, John Legge, Jerry Levy, Henry Liu, Basil Moore, Marc-Andre Pigeon, Clifford Poirot, Jason Potts, Barkley Rosser, Gunnar Tomasson, Sean Toohey, Robert Vienneau, Graham White, and Karl Widerquist, for reading and commenting upon drafts of this book. I

would especially like to thank Karl Widerquist for detailed suggestions on content and the flow of arguments, John Legge for assistance with the proofs of some propositions, Alan Isaac for providing a testing foil to many propositions in the early chapters, and Geoff Fishburn for many years of intelligent and critical discussion of economic theory.

Joyce Hitchings provided valuable feedback on how to make the book's arguments and counter-arguments more accessible to readers with no prior training in economics.

I have also received great encouragement and feedback from my publishers Tony Moore of Pluto Press, and Robert Molteno of Zed Books. My editor, Michael Wall, did a sterling job of making the final product more concise and accessible than the original manuscript.

Sabbatical leave granted by the University of Western Sydney gave me the time away from the everyday demands of an academic life needed to complete a book. The Jerome Levy Institute of Bard College, New York and the Norwegian University of Science and Technology in Trondheim, Norway, kindly accommodated me while the finishing touches were applied to the manuscript.

And so to battle.

1. No more Mr Nice Guy

Why the public needs to know that economics is intellectually unsound

Why have we handed over the running of the world to economists?

It is hardly because economics has won the intellectual equivalent of a popularity poll or election. In fact, if anything, economics is deeply unpopular, and its unpopularity spans all social spectra. The demonstrations against the IMF and other institutions in Seattle, Washington, Davos, Melbourne and Prague were fundamentally demonstrations against the proposition that the world should be reshaped in accordance with economic theory. Those events were simply the most dramatic of many protests which have frequently cut across the standard Left–Right divide of political debate, uniting the most unlikely of bedfellows against the policy recommendations of economists.

Only parties of the middle – the Democrats and Republicans of America, the Social Democrats or Christian Democrats of Europe, England's Conservative Party and 'New Labour', Australia's Liberal and Labor Parties – espouse the policies of economics. Parties which are more avowedly Left and Right often express economic attitudes which are diametrically opposed to policies of centrist parties, but are remarkably similar to each other – with their similarity derived from a shared disdain for conventional economic thought. However, because centrist parties have almost always been in power, the economic policies they champion have shaped the modern world.

At the level of 'grassroot' politics, in countless disputes around the world, social groups which normally oppose each other have also found themselves united by a common opposition to economic policy. Unionists opposing how economics treats labour as a mere commodity have found themselves standing, figuratively or actually, shoulder to shoulder with businessmen chafing against its anti-monopoly dogma. Farmers bemoaning the decline of rural communities have found themselves united with ecologists denying that a dollar value can be placed on nature. Feminists decrying its lack of respect for household labour have found themselves allied with Christians fuming at its portrayal of people as innately hedonistic.

And yet economics has swept all these opponents aside.

Nor is the political success of economics explicable because, though its message may occasionally be unpalatable, its opponents have to concede that it works: that the world has become a clearly better place because of the policies followed by governments that have followed the advice of economists.

The global economy of the early 21st century looks a lot more like the economic textbook ideal than did the world of the 1950s. Barriers to trade have been abolished or dramatically reduced, regulations controlling the flow of capital have been liberalised, currencies are now valued by the market rather than being set by governments; in so many spheres of economic interaction, the government's role has been substantially reduced. All these changes have followed the advice of economists that the unfettered market is the best way to allocate resources, and that well-intentioned interventions which oppose market forces will actually do more harm than good.

With the market so much more in control of the global economy now than fifty years ago, then if economists are right, the world should be a manifestly better place: it should be growing faster, with more stability, and income should go to those who deserve it.

Unfortunately, the world refuses to dance the expected tune. In particular, the final ten years of the 20th century were marked, not by tranquil growth, but by crises: the Japanese economic meltdown, the Long Term Capital Management crisis, the Russian crisis, the Mexican crisis, the Asian crisis, and many more.

Economists are prone to 'point the finger' and blame these crises on particular economic policy failings by the relevant governments – closed capital markets in Japan, fixed exchange rate shenanigans by the Thai government prior to the collapse of the baht, and so on. Yet many non-economists harbour the suspicion that perhaps these crises were in some sense caused by following the advice of economists.

This perspective was recently supported by none other than Joseph Stiglitz, a renowned economist who has had an intimate involvement in public policy via his roles as Chief Economist and Vice-President of the World Bank. Speaking as an insider and an economist, he asserted that these crises were indeed often precipitated by economists.

The most extreme case Stiglitz discusses was the collapse of the Russian economy as Russia attempted to move from a command economy to a market economy in a timescale measured in days rather than years. While he notes that there was a group of eminent economists, including himself and Kenneth Arrow, who favoured a slow transition with an

emphasis upon institutional reform, he says that the day was won by another group

> whose faith in the market was unmatched by an appreciation of the subtleties of its underpinnings. These economists typically had little knowledge of the history or details of the Russian economy and didn't believe they needed any. The great strength, and the ultimate weakness, of the economic doctrines on which they relied is that the doctrines are – or are supposed to be – universal. Institutions, history, or even the distribution of income simply do not matter. Good economists know the universal truths and can look beyond the array of facts and details that obscure these truths. (Stiglitz 2000)

The outcome, as Stiglitz details, was far different from the expectations held by these economists. Rather than enabling Russia to rapidly transmute from moribund socialism to dynamic capitalism,

> The rapid privatization urged upon Moscow by the IMF and the [United States] Treasury Department had allowed a small group of oligarchs to gain control of state assets. While the government lacked the money to pay pensioners, the oligarchs were sending money obtained by stripping assets and selling the country's precious national resources into Cypriot and Swiss bank accounts. (Stiglitz 2000)

In other words, the end result of the IMF's Russian 'capitalist revolution' was not a vibrant, efficient market economy, but a capitalism of crooks – and a drastically impoverished nation. As Stiglitz observes, "standards of living remain far below what they were at the start of the transition. The nation is beset by enormous inequality, and most Russians, embittered by experience, have lost confidence in the free market." Indeed, this free market experiment may have done more to rehabilitate Karl Marx – and even Joseph Stalin – in the eyes of the average Russian, than anything positive done by Russia's socialist rump.

Stiglitz tells a similar tale of the impact economists had on the Asian crisis, where the IMF's enforcement of austerity seriously worsened a crisis which had been initiated by the international capital markets (Stiglitz 1998 and 2000).

Where I and a significant minority of economists part company with Stiglitz is on the explanation he gives to the question he was often asked, of "how smart – even brilliant – people could have created such bad policies". Part of Stiglitz's answer is that "these smart people were not using smart economics". This book puts the case that even the best, latest version of the type of economics Stiglitz describes as smart is not smart, but fundamentally unsound.[3]

Unsmart economics

The belief that economic theory is sound, and that it alone considers 'the big picture', is the major reason why economics has gained such an ascendancy over public policy. Economists, we are told, know what is best for society because economic theory knows how a market economy works, and how it can be made to work better, to everyone's ultimate benefit. Its critics are simply special interest groups, at best misunderstanding the mechanisms of a market economy, at worst pleading their own special case to the detriment of the larger good. If we simply ignore the criticisms, and follow the guidelines of economic policy, ultimately everybody will be better off. The occasional failures of economies to respond as economic theory predicts occur because the relevant policy-makers either applied the theory badly, or were using out-of-date economics.

Bunkum.

If this proposition were true, then economic theory would be clear, unequivocal, unsullied, and empirically verified.

It is nothing of the sort.

Though economists have long believed that their theory constitutes "a body of generalisations whose substantial accuracy and importance are open to question only by the ignorant or the perverse" (Robbins 1932), for over a century economists have shown that economic theory is replete with logical inconsistencies, specious assumptions, errant notions, and predictions contrary to empirical data.

These critical economists are neither 'ignorant' of economic theory, nor 'perverse' in their motives. As this book shows, they have a far more profound understanding of economic theory than those economists who refuse to peer too deeply into the foundations of their dogma. Far from being driven by perversity, they hoped to improve economics by eliminating notions which were illogical, internally inconsistent, or irrelevant to the actual economies in which we live.

When their critiques are collated, little if anything of conventional economic theory remains standing.

Virtually every aspect of conventional economic theory is intellectually unsound; virtually every economic policy recommendation is just as likely to do general harm as it is to lead to the general good. Far from holding the intellectual high ground, economics rests on foundations of quicksand. If

economics were truly a science, then the dominant school of thought in economics would long ago have disappeared from view.

Instead it has been preserved, not via greater knowledge, as its advocates might believe, but by ignorance. Many economists are simply unaware that the foundations of economics have even been disputed, let alone that these critiques have motivated prominent economists to profoundly change their views, and to consequently themselves become, to some extent, critics of economic orthodoxy. Names such as Irving Fisher, John Hicks, Paul Samuelson, Robert Solow, Alan Kirman and Joseph Stiglitz are famous within economics because they made major contributions to modern economic theory. Yet to varying degrees, these and other prominent economists have distanced themselves from conventional economics, after coming to believe, for a range of reasons, that the theory harboured fundamental flaws.

Unfortunately, in a classic illustration of the cliche that 'a little knowledge is a dangerous thing', lesser intellects continue to build the economic edifice atop foundations which many of its architects long ago declared suspect.

There are many reasons for this failure of economics to accept fundamental criticism, and to evolve into a different but richer theory. As I discuss later, these include the undeniable complexity of economic phenomena, and the impossibility of conducting crucial experiments to decide between competing theories. But a key reason – the one which motivated me to write this book – is the manner in which economics is taught.

Educated into ignorance

Most introductory economics textbooks present a sanitised, uncritical rendition of conventional economic theory, and the courses in which these textbooks are used do little to counter this mendacious presentation. Students might learn, for example, that 'externalities' reduce the efficiency of the market mechanism. However, they will not learn that the 'proof' that markets are efficient is itself flawed.

Since this textbook rendition of economics is also profoundly boring, many students do no more than an introductory course in economics, and instead go on to careers in accountancy, finance or management – in which, nonetheless, many continue to harbour the simplistic notions they were taught many years earlier.

The minority which continues on to further academic training is taught the complicated techniques of economic analysis, with little to no discussion of whether these techniques are actually intellectually valid. The enormous critical literature is simply left out of advanced courses, while glaring logical shortcomings are glossed over with specious assumptions. However, most students accept these assumptions because their training leaves them both insufficiently literate and insufficiently numerate.

Modern-day economics students are insufficiently literate because economic education eschews the study of the history of economic thought. Even a passing acquaintance with this literature exposes the reader to critical perspectives on conventional economic theory – but students today receive no such exposure.

They are insufficiently numerate because the material which establishes the intellectual weaknesses of economics is complex. Understanding this literature in its raw form requires an appreciation of some quite difficult areas of mathematics– concepts which require up to two years of undergraduate mathematical training to understand.

Curiously, though economists like to intimidate other social scientists with the mathematical rigour of their discipline, most economists do not have this level of mathematical education.

Instead, most economists learn their mathematics by attending courses in mathematics given by other economists. The argument for this approach – the partially sighted leading the partially sighted – is that generalist mathematics courses don't teach the concepts needed to understand mathematical economics (or the economic version of statistics, known as econometrics). This is quite often true. However, this has the side-effect that economics has produced its own peculiar versions of mathematics and statistics, and has persevered with mathematical methods which professional mathematicians have long ago transcended. This dated version of mathematics shields students from new developments in mathematics that, incidentally, undermine much of economic theory.

One example of this is the way economists have reacted to 'chaos theory' (discussed in Chapter 8). Most economists think that chaos theory has had little or no impact – which is generally true in economics, but not at all true in most other sciences. This is partially because, to understand chaos theory, you have to understand an area of mathematics known as 'ordinary differential equations'.[4] Yet this topic is taught in very few courses on mathematical economics – and where it is taught, it is not covered in sufficient depth. Students may learn some of the basic techniques for handling what are known as 'second-order linear differential equations', but chaos and complexity only begin to manifest themselves in 'third order

nonlinear differential equations'.[5]

Economics students therefore graduate from Masters and PhD programs with an effectively vacuous understanding of economics, no appreciation of the intellectual history of their discipline, and an approach to mathematics which hobbles both their critical understanding of economics, and their ability to appreciate the latest advances in mathematics and other sciences.

A minority of these ill-informed students themselves go on to be academic economists, and then repeat the process. Ignorance is perpetuated.

The attempt to conduct a critical dialogue within the profession of academic economics has therefore failed, not because economics has no flaws, but because – figuratively speaking – conventional economists have no ears. So then, 'No More Mr Nice Guy'. If economists can't be trusted to follow the Queensberry Rules of intellectual debate, then we critics have to step out of the boxing ring and into the streets.

Does economics matter?

Economics is often popularly compared to the weather, for two good reasons. Firstly, just as the climate would exist even if there were no intellectual discipline of meteorology, the economy itself would exist whether or not the intellectual pursuit of economics existed. Unlike something that is consciously constructed, such as an aeroplane, the economy is a product of humanity's evolving systems of production and distribution. We don't need economics in the same sense that we need engineering to design planes, etc. Instead, economics shares a fundamental *raison d'etre* with meteorology, that of attempting to understand a complex system.

Secondly, like weather forecasters, economists frequently get their forecasts of the economic future wrong. But in fact, though weather forecasts are sometimes incorrect, overall meteorologists have an enviable record of accurate prediction – whereas the economic record is tragically bad. Farmers, sailors and ordinary folk who rely on weather forecasts are more often than not beneficiaries of meteorological science. Politicians, businessmen and ordinary folk who rely on economic forecasts are far more often than not misled.

This implies that it would be possible to just ignore economics – to treat it and its practitioners as we these days treat astrologers: as a source of distraction and amusement rather than as a guide to what the future may bring. Unfortunately, there are two ways in which economics differs from meteorology, which mean that this lazy option is not possible.

Firstly, economics, unlike meteorology, is a social discipline. What we believe about economics therefore has an impact upon human society and the way we relate to one another. Its effects upon interpersonal relations matter.

Secondly, while human activity is clearly having a impact on the earth's climate, no meteorologist in her right mind would suggest that we should modify the physical environment to make the climate work better.[6] But economists, despite the abysmal predictive track record of their discipline, are forever recommending ways in which the institutional environment should be altered to make the economy work better.

Because politicians and bureaucrats have believed that economic advice is soundly based, this system-altering advice of economists has been taken seriously and acted upon – despite the frequent opposition of many other segments of society. Thus most of Europe has entered a monetary union that consciously restricts the freedom of its member states to undertake expansionary fiscal policies – because economic theory argues that government deficits are bad. Around the world, governments frequently oppose minimum wage laws – because economists argue that such laws increase unemployment. Tariff barriers have been reduced or eliminated – because economics argues that free trade achieves higher social welfare than regulated trade. Indebted Third World governments have been forced to abandon subsidies on basic commodities – because economists argue that these subsidies reduce economic growth by distorting the price mechanism. Russia was rushed into privatisation – because politically influential economists believed that a rapid transition to capitalism was both possible, and preferable to a gradual transition. And the Western world's middle class has been encouraged to entrust its future security to the ups and downs of the stock market – because economists believe that finance markets are inherently 'efficient'.

Economists would contend that these changes have made the world a better place, not because economists have actually verified that the changes have been beneficial, but because the changes have made the real world look more like the hypothetical world of the economic textbook. Since, in economic models, the hypothetical pure market performs better than the mixed economy in which we live, economists are confident that economic reform makes the world a better place. Where problems have occurred, economists normally assert that this was because their advice was not followed properly.

But this confidence in economic reform begs the question – is the

hypothetical world of the economic textbook actually a better place than the real world, with all its 'distortions'? This is only possible if the economic theory that describes the economist's ideal world is internally consistent. If the theory is internally inconsistent – if it requires impossible conditions to function – then the economic ideal may actually be an entirely useless guide to how the real world works, let alone to how it might be improved. Economic reform could produce a manifestly worse system than the one which it alters.

This book presents the wealth of intellectual evidence which proves that economic theory is internally contradictory. These contradictions are so extreme and pervasive that there is little if any chance that reforms guided by this theory will actually improve the economy or society. Though economists have championed economic reform for the quite altruistic reason that they believe these reforms improve social welfare, their recommendations are far more likely to have made the world a worse place, not a better one.

Thus, the economic conditions imposed to achieve monetary union in Europe could enforce a permanent recession upon Europe, and compromise the ability of its governments to counteract any severe downturn in world economic activity. Trade liberalisation could reduce global economic welfare because the rapid opening up of markets could destroy productive capacity. The abolition of price subsidies could retard economic growth by amplifying class conflict in the highly unequal societies of the Third World. Rapid economic change could lead to social breakdown, rather than the development of vibrant market economies. And America's middle class could find its retirement nest-eggs eliminated by the collapse of a wildly speculative stock market.

Mainstream economists might deride these statements, but as this book shows, their arguments to the contrary are specious.

The public could still afford to ignore economics if the discipline had the ability to correct its own excesses. But it does not. Despite its record at forecasting, despite the evidence that economic theories are not consistent, and despite negative outcomes from 'economic reform', the intellectual discipline of economics shows no tendency to reform itself. Instead, unsound theories continue to be taught to students as if they were incontrovertible. Economics cannot be trusted to reform its own house. Therefore, just as politics is too important to leave to the politicians, economics is too important to leave to the economists.

What is economics?

Economic textbooks usually describe economics as "the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses" (Robbins 1932).

This definition is superficially appealing. However, in the course of this book, I will argue that it is also profoundly misleading. It is used by just two of the many schools of thought which exist within the social science of economics[7] and at least four other schools of thought would disagree with it, for at least the following three reasons:

w The definition takes the existing market economy as given and immutable, or ignores the institutional complexities of economic organisation. However, the capitalist economy evolved out of the previous feudal system, and capitalism continues to evolve and change over time: today's capitalism is a profoundly different creature from the system that Adam Smith championed in 1776, and that Marx condemned in 1867. Similarly, the capitalism of modern Russia is substantially different from the capitalism of the modern USA. Economic theory should consider the processes by which economic systems change, and ways in which institutions affect economic performance. But this definition mitigates against comparative studies, and the examination of evolutionary processes;

w The definition emphasises the static issue of efficiently allocating what we have today, rather than the dynamic issue of how output, demand, employment, etc., change over time. Economists assume that dynamic issues are simply a by-product of achieving static efficiency – that if we get the static efficiency stuff right, then the dynamic problems will be solved too. In fact, static and dynamic issues are often in conflict, so that, for example, something that improves efficiency may in fact reduce growth.[8] Economic theory which teaches us how to achieve static efficiency may therefore be a very bad guide to dealing with a dynamic, changing economy.

w Its focus on the behavioural motivations of individuals ignores the many systemic forces which are generated by and in turn mould individual behaviour. 'Ends' and 'means' are treated as independent in this definition, when in practice commodities, tastes and the means of production interact with and co-determine each other. The process by which consumer demand for 'ends' is produced as new 'means' are invented is a legitimate subject for economic analysis, and yet this definition precludes it by treating 'ends' and 'means' as independent.

I expect that these criticisms of the conventional definition of economics won't really be clear to you until you finish reading this book. However, the fact that some economists object to this definition should indicate that

economics is not quite as monolithic as some economists might think it to be.

In fact, the assertion that only those untrained in economics could object to it is propaganda. There are plenty of economists who have been well trained in conventional economics, and who are even more critical of economics than those who have not had such training. These critical economists are well-informed objectors to the delusion that economics consists of incontrovertible truths.

Clearly then, my target in this book is not economics in general, but the dominant school of thought within economics. This school is technically known as neoclassical economics, though for the sake of brevity I will normally refer to it as simply 'economics' (though in some contexts I will describe it as 'mainstream', 'conventional' or 'orthodox' economics). It will receive the lion's share of my critical attention in this book – though occasionally some of the minority schools (described in Chapter 14) will come in for criticism.

However, in general, the minority schools will be put forward to show how economics could be rescued from its current parlous state of dangerous irrelevance. This is not only because this gives the lie to the claim by economists that economic reform – by which they mean removing aspects of the real world which don't conform to mainstream economic theory – is inevitable and ultimately in everyone's best interests. It is also because, since mainstream economics is profoundly wrong, the existence of alternatives to it is highly important. The rejection of the dominant approach to economics does not leave us in an intellectual vacuum.

Is reform of economics possible?

If economics today is so bad, is there any hope of ever understanding the complex social system in which we live? I would not have written this book if I did not believe that the answer was, at least tentatively, 'yes'. The critiques collated in this book will, I hope, convince you that the dominant school of thought in economics is useless as a guide to understanding a capitalist economy. However, the minority schools of thought have between them the beginnings of an appreciation of this complex, unstable, but dynamic social system.

If the hegemony of conventional thinking can be broken, then there is some chance that, as humanity begins its fifth century of sailing on a capitalist sea, economists might finally construct a realistic map of this troubled ocean. This will not mean that the squalls that periodically beset it can be avoided. But at least in future we may not sail the ship full speed into icebergs.

Debunking Economics: a user's guide

Who is this book for?

Interest in economics as an intellectual pursuit for its own sake has waned significantly over the last twenty years, and I have often heard academic economists lament this fact – especially since falling student enrolments have undermined their job security.

I am not at all amazed by this drop in interest: it is a predictable side-effect of the very philosophy of life which economists espouse. They have told all and sundry that the world would be a better place if we all focussed upon our own self-interest, and let the market take care of the common good. Why then is it surprising that students have swallowed this spiel, and decided to study subjects which more clearly lead to a well-paid job – business management, human resources, computing, etc. – rather than to study economics?

This book has been written precisely for that audience which economists once derided, and whose absence they now lament: people who are interested in 'the common good'. Its message, that the economic mantra ('individuals should pursue their own interests and leave society's overall interests to the market') is wrong, is not new. Many books have made the same point in the past. What is new about this book is that it makes that point using economic theory itself.

This book should also be useful to budding students of economics, in at least two ways. Firstly, unless they are lucky enough to attend some of the few universities where pluralism rules, they are about to submit to an education in economics that is in reality an indoctrination. This book covers the issues which should form part of an education in economics, but which are omitted by the vast majority of textbooks.

Secondly, they should find that the explanations of economic theory in this book make it easier to pass exams in economics. I have found that one of the main barriers which new students face in learning economics sufficiently well to be able to pass exams in it, is that they can't reconcile the theory with their own 'gut feelings' about economic issues. Once students realise that they should trust their gut feelings, and treat economic theory as irrelevant to the real economy, then suddenly it becomes much easier to pass exams. Just treat economics like a game of chess, play the games the exam questions require of you, and you'll pass easily (just don't mention the inconsistencies in the rules!).

If you are already a somewhat uncomfortable student of economics, but you lack confidence because you are surrounded by peers who can't understand your disquiet, then this book should allay your fears. Normally, the journey from a troubled student to an informed critic is a difficult and lonely one. I hope to make that journey far less difficult, and less lonely.

Similarly, I hope that professional critical economists will find this book a useful introductory compendium to those many critiques of economic theory that are currently scattered through dozens of books and hundreds of journal articles. While the arguments are not presented with the rigour of those formal critiques, the book provides an accessible and understandable introduction to that important and neglected literature. The curious student can be told to use this book as a guide before delving into the more difficult, formal literature.

Because it explains and debunks economic theory from first principles, this book will also be of use to anyone whose career makes them reliant upon the advice of economists. Hopefully it will encourage such people to look more widely for advice in future.

Last but by no means least, this book provides the mainstream economist with a compendium of all the critical topics which should form a part of advanced education in economics. I can guarantee that mainstream economists will hate the irreverent tone of this book. Nonetheless I'd ask them to persevere with open – but sceptical – minds. I hope that exposure to the many published critiques of economics might encourage them to take a good, hard, dispassionate look at a theory which I believe they have accepted too uncritically.

What's in this book?

This book has been primarily written for people who are inclined to be critical of economics, but who are intimidated by its apparently impressive intellectual arsenal. I start from the premise that, though you might be familiar with the conclusions of economic theory, you are unfamiliar with how those conclusions were derived. You therefore don't have to have studied economics previously to be able to read this book.

I have also eschewed the use of mathematical formulas.[9] Though I frequently use mathematics in my own research, I'm well aware of the impact that mathematical symbols have on the intelligent lay reader (a Norwegian colleague calls it the MEGO effect: 'My Eyes Glaze Over'). Instead, where some mathematical concept is needed to understand a critique, I present it, as well as is possible, in verbal, tabular, or diagrammatic form (though as noted below, mathematical argument is available, for those who want to see it, at the book's website).

Despite the absence of mathematics, this book will still require significant intellectual exertion by the reader. The arguments of economic theory are superficially appealing, as Veblen long ago observed. To understand why they are nonetheless flawed requires thought at a deeper level than just that of surface appearances. I have attempted to make both economic theory and the flaws behind it relatively easy to comprehend, but there will be times when the difficulty of the material defeats my abilities as an expositor.

This problem is amplified by the fact that this book is effectively two books in one.

Firstly, it provides a detailed exposition of the conventional theory, and takes none of the short-cuts taken by the vast majority of conventional economic texts. As I noted above, one reason why economic instruction takes short-cuts is because the foundations of conventional economics are not only difficult to grasp, but also profoundly boring. Economics should be an exciting, stimulating intellectual challenge, but conventional economics almost goes out of its way to be mundane. Unfortunately, I have to explain conventional economics in detail in order to be able to discuss the critiques of this theory. There are thus sections of this book which are inherently tedious – despite my attempts to lighten the discourse. This applies especially to the chapters on the neoclassical theories of consumption (Chapter 2) and of production (Chapter 3).

Secondly, this book provides a detailed debunking of conventional theory. This is, I hope, rather more interesting than conventional theory itself – though nowhere near as interesting as an exposition of a truly relevant economics would be. But it is quite possible that the exposition of conventional theory which precedes each debunking may persuade you that the conventional economic argument makes sense. Your mind will therefore be tossed first one way and then the other, as you first grind through understanding the foundations of conventional economics, and then attempt to comprehend profound but subtle critiques of the superficially convincing conventional logic.

So, especially if you have never read a book on economic theory, you will undoubtedly find some sections very difficult. You may therefore find it easier to treat this book as a reference work, by reading Part One (chapters 2–5) carefully, and then turning to the rest when you have some specific economic issue to explore.

Alternately, you can read the chapters in Parts Two (chapters 6–10) and Three (chapters 11–14) before you attempt the earlier, foundation ones.

This is possible because in these later chapters I 'cut economics some slack', and accept concepts which have in fact been debunked in the earlier chapters. After you've considered the failings of economics in these more interesting applied areas, you could then turn to the flaws in its foundations.

Whichever way you approach it, this book will be a difficult read. But if you are currently a sceptic of economics, and you wish to develop a deeper understanding of why you should be sceptical, I believe the effort will be worth it.

Not Left versus Right but right versus wrong

One possible interpretation of this book – certainly one I expect to get from many economists – is that it is just a left-wing diatribe against rational economics. This common response to intellectual criticism – categorise it and then dismiss it out of hand – is one of the great sources of weakness in economics, and indeed much political debate.

It is probably true that the majority of those who criticise conventional economic theory are closer to the left than the right end of the political spectrum – though as I note above, there are many profoundly right-wing critics of conventional economics. Only parties occupying the middle of the political spectrum tend to espouse and implement conventional economics.

However, the critiques in this book are not based on politics, but on logic. No political position – left, right or middle – should be based on foundations which can easily be shown to be illogical. Yet much of conventional economic theory is illogical. Those who occupy the centre stage of modern politics should find a firmer foundation for their politics than an illogical economic theory.

The same comment, of course, applies to those at the left-wing end of the political spectrum, who base their support for radical social change on conventional Marxian economics. As I argue in Chapter 13, conventional Marxism is as replete with logical errors as is neoclassical economics.

Escher without the panache

One thing which sets economics apart from other social sciences, and which makes it hard for non-economists to understand economics, is the extent to which its arguments are presented in the form of diagrams. Even leading economists, who develop their theories using mathematics, will often imagine their models in diagrammatic form.

These diagrams represent models which are supposed to be simplified but nonetheless accurate renditions of aspects of the real-world phenomena of production, distribution, exchange, consumption, and so on. When an economist talks of the economy behaving in a particular fashion, what she really means is that a model of the economy – and normally a graphical model – has those characteristics.

To learn economics then, one has to learn how to read diagrams and interpret the models they represent. This applies to critics as much as believers, but the very act of learning the diagrams tends to separate one from the other. Most critical thinkers find the process tedious, and drop out of university courses in economics. Most of those who stay behind become seduced by the diagrams and models, to the point where they have a hard time distinguishing their models from reality.

The critical thinkers, who could not cope with the diagrammatic representation of economic reality, were fundamentally correct: economic reality cannot be shoehorned into diagrams. Consequently, these diagrams often contain outright fallacies, conveniently disguised by smooth curves.

In other words, rather than being accurate renditions of the economy, the standard economic diagrams are rather like Escher drawings, in which the rules of perspective are used to render scenes which appear genuine – but which are clearly impossible in the real, three-dimensional world.

Whereas Escher amused and inspired with his endless staircases, eternal waterfalls and the like, economists believe that their models give meaningful insights into the real world. But they could only do so if the Escher-like assumptions economists make could apply in reality – if, metaphorically speaking, water could flow uphill. Since it cannot, economic models are dangerously misleading when used to determine real-world policy.

Obviously therefore, I do not wish to encourage you to 'think diagrammatically', since this mode of thought has helped to confuse economics rather than to inform it. However, to be able to understand where economics has gone wrong, you need to see what has led it astray. So while I will attempt as much as possible to supplement the conventional diagrammatic explanations of models with verbal and tabular expositions, it is still likely that to be able to fully comprehend this book you will need to learn how to read diagrams – though not, I hope, to believe them.

Blow by blow

In most chapters, I take a key facet of economics, and first state the theory as it is believed by its adherents. I then point out the flaws in this superficially appealing theory – flaws that have been established by economists and, in most instances, published in economic journals. As I show, the effect of each flaw is normally to invalidate the theoretical point completely, yet in virtually every case, economics continues on as if the

critique had never been made.

Economics is a moving target, and the outer edges of the theory sometimes bear little resemblance to what is taught at undergraduate level. I concentrate upon the fare served up to undergraduates, rather than the rarefied extremities of new research – mainly because this is the level at which most economists operate, but also because much of the work done at the theoretical 'cutting edge' takes as sound the foundations learnt during undergraduate days.

Chapter by chapter

Part One, Foundations, considers issues which form part of a standard education in economics – the theories of demand, supply, and income distribution – and shows that these concepts have very rickety foundations. It has four chapters:

w Chapter 2 (*The calculus of hedonism*) reveals that economics has failed to derive a coherent theory of consumer demand from its premise that people are no more than self-interested hedonists. As a result, economic theory can't justify a crucial and seemingly innocuous element of its analysis of markets, that demand for a product will fall smoothly as its price rises. A less esoteric aspect of this failure is that economic theory can't justify the standard economic proposition that prices set by an unfettered market system are those which maximise social welfare.

w Chapter 3 (*The price of everything and the value of nothing*) argues that the theory of supply is also flawed, because the conditions which are needed to make the theory work are unlikely to apply in practice. The concept of diminishing marginal returns, which is essential to the theory, is unlikely to apply in practice, 'supply curves' are likely to be flat, or even downward-sloping, and the economic rule for maximising profit does not in fact maximise profits.

w Chapter 4 (*Size does matter*) shows that the economic theory of 'the firm' is logically inconsistent. When the inconsistencies are removed, economic theory cannot distinguish between competitive firms and monopolies, despite its manifest preference for small competitive firms over large ones.

w Chapter 5 (*To each according to his contribution*) looks at the theory of the labour market. The theory essentially argues that wages in a market economy reflect workers' contributions to production. Flaws in the underlying theory imply that wages are not in fact based on merit, and that measures which economists argue would reduce unemployment may in fact increase it.

Part Two, Complexities, considers issues which should be part of an education in economics, but which are either omitted entirely or trivialised in standard economics degrees. It has five chapters:

w Chapter 6 (*The holy war over capital*) complements Chapter 5 by showing that the theory of capital is logically inconsistent. Profit does not reflect capital's contribution to output, and changing the price of capital relative to labour may have 'perverse' impacts on demand for these 'factors of production'.

w Chapter 7 (*There is madness in their method*) examines methodology and finds that, contrary to what economists tell their students, assumptions do matter.

w Chapter 8 (*Let's do the Time Warp Again*), discusses the validity of applying static (timeless) analysis to economics when the economy is clearly dynamic itself. The chapter argues that static economic analysis is invalid when applied to a dynamic economy, so that economic policy derived from static economic reasoning is likely to harm rather than help an actual economy.

w Chapter 9 (*The sum of the parts*) tracks the development of macroeconomics into its current sorry state, and argues that what has been derided as 'Keynesian' macroeconomics was in fact a travesty of Keynes's views.

w Chapter 10 (*The Price is Not Right*) deals with the economic theory of asset markets, known as the 'efficient markets hypothesis'. It argues that the conditions needed to ensure what economists call market efficiency – which include that investors have identical, accurate expectations of the future, and equal access to unlimited credit – cannot possibly apply in the real world. Finance markets cannot be efficient, and finance does affect the real economy.

Part Three, Alternatives, briefly considers alternative approaches to economics. It has four chapters:

w Chapter 11 (*Finance and economic breakdown*) presents three non-equilibrium approaches to the analysis of asset markets, all of which indicate that finance can destabilise the real economy.

w Chapter 12 (*Don't Shoot Me, I'm Only the Piano*) examines the role of mathematics in economic theory. It argues that mathematics itself is not to blame for the state of economics today, but instead that bad and inappropriate mathematics by economists has resulted in them persisting with an inappropriate static equilibrium analysis of the economy. The dynamic, non-equilibrium social system that is a market economy should be analysed with dynamic, non-equilibrium tools.

w Chapter 13 (*Nothing to lose but their minds*) dissects Marxian economics, arguing that this potential alternative to conventional economics is seriously flawed. However, much of the problem stems from an inadequate understanding of Marx by not just his critics, but also his alleged friends.

w Finally, Chapter 14 (*There are alternatives*), briefly presents several alternative schools in economics, and shows that viable if somewhat under-developed alternative ways to 'think economically' already exist.

There's even more on the web

This book does not begin and end with the chapters just mentioned. It is also intimately linked to one of my two websites, http://www.debunkingeconomics.com (my other website, http://www.debtdeflation.com, currently supports my blog on the financial crisis and ultimately will be the online companion to my next book *Finance and Economic Breakdown*.

The website complements the book in several ways. Firstly, sections of the argument have been placed on the web. These are technically necessary, but somewhat tedious, and therefore could distract attention from key issues. Examples include the concepts of 'revealed preference' and 'compensated demand curves' in Chapter 2. These web entries are noted in the text with a comment like "I've skipped explaining a concept called revealed preference. Check the link More/Hedonism/Revealed if you want the full version", which indicates both what has been placed on the web, and where it is located.

Secondly, more lengthy discussion of some topics has been placed on the web. For instance, the failure of the conventional theory of market demand means that alternative approaches must be developed. These, and additional critiques of conventional theory are on the website and referred to under the heading 'But wait, there's more'. The locations of these additional discussions are given by comments like "These and other issues are discussed on the web. Follow the links to More/Hedonism." These sections raise many issues which should be of interest to those critical of conventional economics.

Thirdly, while there are no mathematical formulas used in this book, the logic underlying many of the critiques is mathematical. The mathematically inclined reader can check the original logic by consulting the website. These links are indicated by a parenthetical statement such as "(follow the link Maths/Size/PC_eq_M for the maths)".

Fourthly, some related topics are not covered in the book. One obvious omission is the theory of international trade. The major reason for this omission is that, while sound critiques of international trade theory exist, what I regard as the most obvious and telling critique has not yet been formally developed (I outline this on the website at the link Trade, as well as discussing the formal critiques that have been published). Another reason is that the theory of international trade also depends on many basic concepts that are thoroughly debunked in this book.

Passing judgment on modern economics

This book can be thought of as a critical report card on economics at the beginning of the third millennium. Economic theory, as we know it today, was born in the late 19th century in the work of Jevons, Walras, Menger and (somewhat later) Marshall. I have a reasonably high regard for these founders of what has become mainstream economics. They were pioneers in a new way of thinking, and yet, in contrast to their modern disciples, they were often aware of possible limitations of the theory they were trying to construct. They expected their heirs to extend the boundaries of economic analysis, and they expected economics to develop from the precocious but hobbled child to which they gave birth into a vibrant and flexible adult.

Instead, economics today is ridden with internal inconsistencies: an economic model will start with some key proposition, and then contradict that proposition at a later stage. For example, the theory of consumer demand begins with the proposition that each consumer is unique, but then reaches a logical impasse which it sidesteps by assuming that all consumers are identical.

This raises an important general point about scientific theories. Any theory will have some starting point, which can be chosen in any of a number of ways. Newtonian physics, for example, began with the starting point that any object subject to a force (in a vacuum) will accelerate; Einsteinian physics began with the starting point that the speed of light (also in a vacuum) sets an absolute speed limit for any material object.

Clearly the starting point of a theory can be challenged, but the basis of such a critique is normally what we might term 'external consistency'. That is, since the theory is supposed to describe some objective reality, it must be possible to show significant consistency between the predictions of the theory and that objective reality.

Here the degree of proof often comes down to some statistical measure of accuracy. Using the example of physics again, it is obvious that, at the speeds which humans could impart to a physical body during the 19th

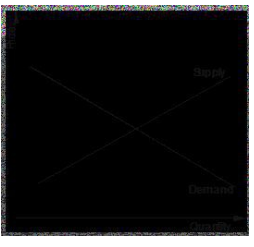
century, the Newtonian vision was extremely accurate.

Internal consistency, on the other hand, requires that everything within the theory must legitimately follow from its starting point. Here, statistical accuracy is not good enough: the fit of the theory with the starting point from which it is derived must be exact. If a theory at some point requires a condition which contradicts its starting point, or any other aspect of itself, then the theory is internally inconsistent and therefore invalid. It is possible to criticise much of economics on the basis that 'reality isn't like that' – and this is occasionally done in the subsequent chapters. However, in general I take two allegedly related aspects of economic theory – the theory of individual consumption and the theory of the market demand curve, for example – and show that to get from one to the other, a clearly contradictory condition must be imposed.

Atheory cannot survive with such contradictions – or rather, it should not. They are clear signals that something is fundamentally wrong with the starting position of the theory itself, and that real progress involves radically revising or even abandoning that starting point. Even some of the most committed economists have conceded that, if economics is to become less of a religion and more of a science, then the foundations of economics should be torn down and replaced. However, if left to its own devices, there is little doubt that the profession of academic economics would continue to build an apparently grand edifice upon rotten foundations.

The founding fathers of modern economics would, I expect, be surprised to find that a manner of thinking they thought would be transitional has instead become ossified as the only way one can do economics and be respectable. They would, I hope, be horrified to find that the limitations of economic theory have been soundly established, and that most 'respectable' economists nevertheless transgress these limits without conscience, and often without knowledge.

Respectability be damned. Like the populace watching the parade of the emperor, respectability has led us to kowtow to a monarch in fine cloth, when an unindoctrinated child can see that the emperor has no clothes. It's time to expose the nakedness of neoclassical economics.



Part 1
Foundations
The logical flaws in the key concepts of conventional economics

The belief that price and quantity are jointly determined by the interaction of supply and demand is perhaps the most central tenet of conventional economics. In Alfred Marshall's words, supply and demand are like the two blades of a pair of scissors: both are needed to do the job, and it's impossible to say that one or the other determines anything on its own. Demand for a commodity falls as its price rises, supply rises as price rises, and the intersection of the two curves determines both the quantity sold and the price.

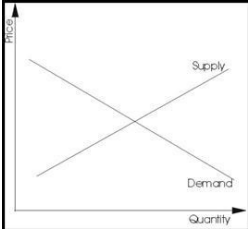


Figure 1: Supply and demand jointly determine price and quantity

This argument still forms the core of modern instruction in economics, and much of economic policy is directed at allowing these twin determinants to act free and unfettered, so that economic efficiency can be at its maximum. But both mainstream and dissident economists have shown that the real world is not nearly so straightforward as Marshall's famous analogy. The next three chapters show that the 'blades of supply and demand' cannot work in the way economists believe.

2. The calculus of hedonism

Why the pursuit of individual self-interest does not maximise social welfare

The corollary to Maggie Thatcher's famous epithet that "There is no such thing as society" is The Life of Brian's "You are all individuals." Maggie's masterpiece succinctly expresses the economic theory that the best social outcomes result from all individuals looking after their own self-interest: the market will ensure that the welfare of all is maximised.

This hedonistic, individualistic approach to analysing society is a source of much of the popular opposition to economics. Surely, say the critics, people are more than just self-interested hedonists, and society is more than just the sum of the individuals in it?

Economists will concede that their model does abstract from some of the subtler aspects of humanity and society. However, they assert that treating individuals as self-interested hedonists captures the essence of their economic behaviour, while the collective economic behaviour of society can be derived by summing the behaviour of this self-interested multitude. The belief that the economic aspect of society is substantially more than the sum of its parts, they say, is misguided.

This is not true. Though mainstream economics began by assuming that this hedonistic, individualistic approach to analysing consumer demand was intellectually sound, it ended up proving that it was not. The critics were right: society is more than the sum of its individuals members, and a society's behaviour cannot be modelled by simply adding up the behaviours of all the individuals in it. To see why the critics have been vindicated by economists, and yet economists still pretend that they won the argument, we have to take a trip down memory lane to late 18th century England.

The kernel

Adam Smith's famous metaphor that a self-motivated individual is led by an 'invisible hand' to promote society's welfare asserts that self-centred behaviour by individuals necessarily leads to the highest possible level of welfare for society as a whole.[10] Modern economic theory has attempted, unsuccessfully, to prove this assertion. The attempted proof had several components, and in this chapter we check out the component which models how consumers decide which commodities to purchase.

According to economic theory, each consumer attempts to get the highest level of satisfaction she can from her income, and she does this by picking the combination of commodities she can afford which gives her the greatest personal pleasure. The economic model of how each individual does this is intellectually watertight (though there are good reasons for objecting to the proposition that people are motivated solely by hedonism).

However, economists encountered fundamental difficulties in moving from the analysis of a solitary individual to the analysis of society, because they had to 'add up' the pleasure which consuming commodities gave to different individuals. But personal satisfaction is clearly a subjective thing, and there is no objective means by which one person's satisfaction can be added to another's. Any two people get different levels of satisfaction from consuming, for example, an extra banana, so that a change in the distribution of income which effectively took a banana from one person and gave it to another could result in a different level of social well-being.

Economists were therefore unable to prove their assertion, unless they could somehow show that altering the distribution of income did not alter social welfare. They worked out that two conditions were necessary for this to be true: (a) that all people have to have the same tastes; (b) that each person's tastes remain the same as her income changes, so that every additional dollar of income was spent exactly the same way as all previous dollars – for example, 20 cents per dollar on pizza, 10 cents per dollar on bananas, 40 cents per dollar on housing, etc.

The first assumption in fact amounts to assuming that there is only one person in society (or that society consists of a multitude of identical drones) – since how else could 'everybody' have the same tastes? The second amounts to assuming that there is only one commodity – since otherwise spending patterns would necessarily change as income rose.

These 'assumptions' clearly contradict the case economists were trying to prove, since they are necessarily violated in the real world. Economic theory therefore undermines Adam Smith's invisible hand, by proving that it is impossible to fulfil the conditions needed to guarantee this outcome. Sadly, however, this is not how most economists have interpreted these results.

When conditions (a) and (b) are violated, as they must be in the real world, then several important concepts which are important to economists collapse. The key casualty here is the vision of demand for any product falling as its price rises. Economists can prove that 'the demand curve slopes downward in price' for a single individual and a single commodity. But in a society consisting of many different individuals with many different

commodities, the 'market demand curve' is more probably jagged, and slopes every which way. One essential building block of the economic analysis of markets, the demand curve, therefore does not have the characteristics needed for economic theory to be internally consistent.

The roadmap

The chapter opens with an outline of Jeremy Bentham's philosophy of utilitarianism, which is the philosophical foundation for the economic analysis of individual behaviour. The conventional economic analysis is outlined through a sequence of tables and graphs of increasing complexity. The presentation highlights the chapter's punchline: that economic theory cannot derive a coherent analysis of market demand from its watertight but ponderous analysis of individual behaviour.

Pleasure and pain

The true father of the proposition that people are motivated solely by self-interest is not Adam Smith, as is often believed, but his contemporary, Jeremy Bentham. With his philosophy of 'utilitarianism', Bentham explained human behaviour as the product of innate drives to seek pleasure and avoid pain. Bentham's cardinal proposition was that

> Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne. They govern us in all that we do, in all we say, in all we think; every effort we can make to throw off our subjection, will serve but to demonstrate and confirm it. In a word a man may pretend to abjure their empire; but in reality he will remain subject to it all the while. (Bentham 1780)

Thus Bentham saw the pursuit of pleasure and the avoidance of pain as the underlying causes of everything done by humans, and phenomena such as a sense of right and wrong as merely the surface manifestations of this deeper power. You may do what you do superficially because you believe it to be right, but fundamentally you do it because it is the best strategy to gain pleasure and avoid pain. Similarly, when you refrain from other actions because you say they are immoral, you in reality mean that, for you, they lead to more pain than pleasure.

Today, economists similarly believe that they are modelling the deepest determinants of individual behaviour, while their critics are merely operating at the level of surface phenomena. Behind apparent altruism, behind apparent selfless behaviour, behind religious commitment, lies self-interested individualism.

Bentham called his philosophy the 'principle of utility' (Bentham 1780), and he applied it to the community as well as the individual. Like his Tory disciple some two centuries later, Bentham reduced society to a sum of individuals:

> The community is a fictitious body, composed of the individual persons who are considered as constituting as it were its members. The interests of the community then is, what? – the sum of the interests of the several members who compose it. It is in vain to talk of the interest of the community, without understanding what is in the interest of the individual. (Bentham 1780)

The interests of the community are simply the sum of the interests of the individuals who comprise it, and Bentham perceived no difficulty in performing this summation:

> An action then may be said to be conformable to the principle of utilitywhen the tendency it has to augment the happiness of the community is greater than any it has to diminish it. (Bentham 1780)

This last statement implies measurement, and Bentham was quite confident that individual pleasure and pain could be objectively measured, and in turn summed to divine the best course of collective action for that collection of individuals called society.[11] Bentham's attempts at such measurement look quaint indeed from a modern perspective, but from this quaint beginning economics has erected its complex mathematical model of human behaviour. Economists use this model to explain everything from individual behaviour, to market demand, to the representation of the interests of the entire community. However, as we shall shortly see, economists have shown that the model's validity terminates at the level of the single, solitary individual.

Flaws in the glass

In most chapters, I attempt to separate the exposition of a concept from its critique. This is simply not possible in this first substantive chapter. Non-economists may criticise the economic representation of a human being as a totally self-interested entity, but they probably expect this economic representation to be internally consistent.

It is not. While economics can provide a coherent analysis of the individual in its own terms, it is unable to extrapolate this to an analysis of

society as the simple sum of its individual members. Unlike many other critiques of economic theory, most mainstream academic economists are aware of this problem, but they pretend that the failure can be managed with a couple of assumptions. Yet the assumptions themselves are so absurd that only someone with a grossly distorted sense of logic could accept them. That grossly distorted sense of logic is acquired in the course of a standard education in economics.

This 'education' begins with students being taught conclusions which would apply if the theory had no logical flaws. Students normally accept that these conclusions have been soundly derived from the basic economic propositions of individual behaviour, and they are in no position to believe otherwise, since the basic building blocks of this analysis are not taught at the introductory level because they are 'too hard'.

This abbreviated induction is sufficiently boring to dissuade the majority of business students from pursuing further economics, and they graduate in some other discipline. However, a minority find the game intriguing, and continue on to another year.

In later undergraduate years, they finally encounter indifference curves, the derivation of the individual demand curve, and the representation of society's interests in the form of a 'social indifference map'. The mildly relevant 'Engels curve' and the complete chimera of the 'Giffen good' are explored as apparent applications of the theory. Market demand curves, and sometimes the basic concepts of 'general equilibrium' (the conditions under which many markets will simultaneously be in equilibrium) are discussed – again with the presumption that the step from the individual to the aggregate is valid.

Most economics graduates seek employment in the private sector, and parts of the public sector, where they normally champion the economics perspective. However, a minority of this minority pursues further study, to seek employment as academic economists – and in search of education rather than remuneration, since private and even public sector salaries are far higher than academic ones. Once they have embarked upon this road to ordination as an economist, most students are fully inculcated with the mainstream economic way of thinking.

Finally, in Honours, Masters or PhD courses, they study the full exposition given below, and finally learn that the aggregation of individual demand is only valid under patently absurd conditions. However, by this time the indoctrination into the economic mind set is so complete that most of them cannot see the absurdity. Instead, they accept these conditions as no more than simple devices to sidestep pesky but minor problems, so that 'rational' economic analysis can be undertaken.

It would be easy to accede to a simplistic conspiracy theory to explain why economic education takes such a convoluted route on this issue. However, I believe the explanation is both more mundane and more profound.

At the mundane level, the proposition that individual behaviour is motivated by utility maximisation, the concept of a downward-sloping demand curve, and the vision of society as simply an aggregate of individuals, are easier to grasp than the many qualifications which must be applied to keep these notions intact. Academic economists therefore instruct their students in the easy bits first, leaving the difficult grist for higher-level courses.

At the profound level, it reflects the extent to which economists are so committed to their preferred methodology that they ignore or trivialise points at which their analysis has fundamental weaknesses. Were economics truly worthy of the moniker 'social science' these failures would be reason to abandon the methodology and search for something sounder.

Whatever the reasons, this lazy pedagogy bifurcates economics students into three camps. The vast majority study a minimum of economics in a business degree and graduate unaware of any flaws in the glass. Members of the second, much smaller group go on to professional academic careers, and treat the flaws as marks of a fine crystal, rather than clear evidence of a broken vessel. The third, a handful, become critics within the profession, who aspire to build more realistic theories and, sometimes, try to make the second group see the cracks in their beloved but broken goblet.

These sentiments may appear extreme now, but I doubt that they will appear so by the time you have read this chapter.

Now pour yourself a cup of coffee – or any other appropriate stimulant. The next section is crucial to understanding both economic theory and its weaknesses, but it can't help but be boring.

"The sum of the interests"

Bentham's statement that "The community is a fictitious body . . . The interests of the community then is (sic)the sum of the interests of the several members who compose it" is no more than an assertion. To turn this into a theory, economists had to achieve two tasks: to express Bentham's analysis mathematically, and to establish mathematically that it was possible to derive social utility by aggregating individual utility.

One century after Bentham, the founders of neoclassical economics accomplished the first task with relative ease. Over time, the representation of these concepts matured from simple but flawed notions to arcane but watertight models of individual behaviour.

The individual consumer as represented by economic theory

In keeping with the notion that beneath all individual actions lie the motivations of pleasure seeking and pain avoidance, early attempts to use utility theory to explain behaviour – by which economists meant almost exclusively the consumption of commodities[12] – postulated that each unit consumed of any commodity yielded a certain number of underlying units of satisfaction, called 'utils'. Additional units of a given commodity resulted in a smaller number of additional utils. The picture is as shown in Table 1 and Figure 1.

Bananas	Utils	Change in Utils
1	8	8
2	15	7
3	19	4
4	20	1

Table 1: 'Utils' and change in utils from consuming bananas

Here, one unit of a commodity – say, a banana – yields 8 'utils' of satisfaction to the consumer. Two bananas yield 15 utils, so that the second banana has contributed seven additional utils to the consumer's satisfaction: one less than the first banana, but still a positive quantity. Three bananas yields 19 utils, so that the change in utils from consuming the third banana is 4 utils.

Figure 1 reproduces this table in graphical format. The first, darker set of bars shows total utility, while change in utility is shown by the second set of bars. Total utility always rises (hedonists are assumed to always prefer more to less), but at a decreasing rate.

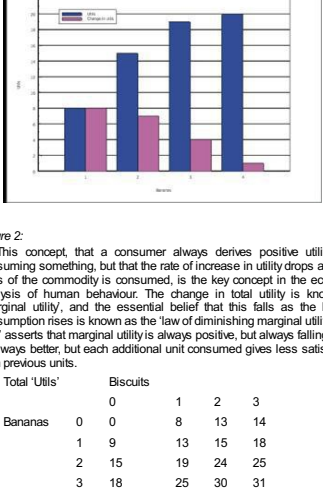


Figure 2:

This concept, that a consumer always derives positive utility from consuming something, but that the rate of increase in utility drops as more units of the commodity is consumed, is the key concept in the economic analysis of human behaviour. The change in total utility is known as 'marginal utility', and the essential belief that this falls as the level of consumption rises is known as the 'law of diminishing marginal utility'. This 'law' asserts that marginal utility is always positive, but always falling: more is always better, but each additional unit consumed gives less satisfaction than previous units.

Total 'Utils'		Biscuits			
		0	1	2	3
Bananas	0	0	8	13	14
	1	9	13	15	18
	2	15	19	24	25
	3	18	25	30	31

Table 2: Utils arising from the consumption of two commodities

Obviously, utility is derived from consuming more than just one commodity. Economists assume that the law of diminishing marginal utility applies across all commodities, so that additional units of any commodity give the consumer positive but falling amounts of utility. This can be illustrated by having two commodities on the one graph, with the actual level of utility they generate shown as the height on a '3D' graph.

This is shown in Table 2 and Figure 2, where the first commodity is bananas, and the second, biscuits. Each number in the table shows how

many utils the consumer garnered from each combination of bananas and biscuits. Graphically, this yields a set of 3D bars, with the bars getting ever higher as more biscuits and bananas are consumed.

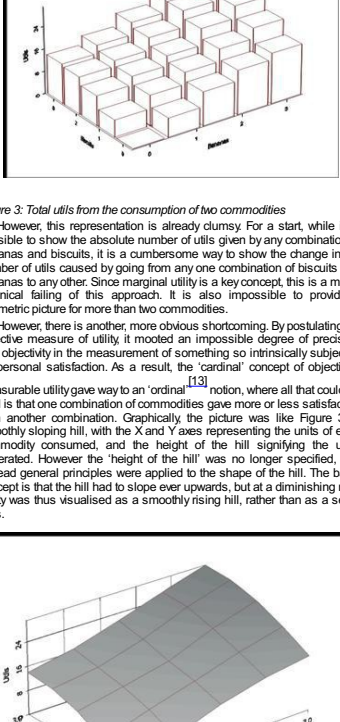


Figure 3: Total utils from the consumption of two commodities

However, this representation is already clumsy. For a start, while it is possible to show the absolute number of utils given by any combination of bananas and biscuits, it is a cumbersome way to show the change in the number of utils caused by going from any one combination of biscuits and bananas to any other. Since marginal utility is a key concept, this is a major technical failing of this approach. It is also impossible to provide a geometric picture for more than two commodities.

However, there is another, more obvious shortcoming. By postulating an objective measure of utility, it mooted an impossible degree of precision and objectivity in the measurement of something so intrinsically subjective as personal satisfaction. As a result, the 'cardinal' concept of objectively measurable utility gave way to an 'ordinal'[13] notion, where all that could be said is that one combination of commodities gave more or less satisfaction than another combination. Graphically, the picture was like Figure 3: a smoothly sloping hill, with the X and Y axes representing the units of each commodity consumed, and the height of the hill signifying the utility generated. However the 'height of the hill' was no longer specified, and instead general principles were applied to the shape of the hill. The basic concept is that the hill had to slope ever upwards, but at a diminishing rate. Utility was thus visualised as a smoothly rising hill, rather than as a set of bars.

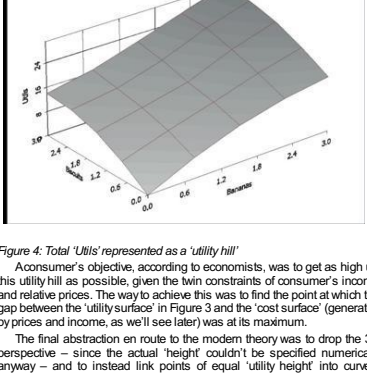


Figure 4: Total 'Utils' represented as a 'utility hill'

A consumer's objective, according to economists, was to get as high up this utility hill as possible, given the twin constraints of consumer's income and relative prices. The way to achieve this was to find the point at which the gap between the 'utility surface' in Figure 3 and the 'cost surface' (generated by prices and income, as we'll see later) was at its maximum.

The final abstraction en route to the modern theory was to drop the 3D perspective – since the actual 'height' couldn't be specified numerically anyway – and to instead link points of equal 'utility height' into curves. Topographically, these curves were equivalent to the contours on a geographic map indicating locations of equal height, or isobars on a weather chart indicating regions of equal pressure. Using the same data as in Table 2, this yields Figure 4.

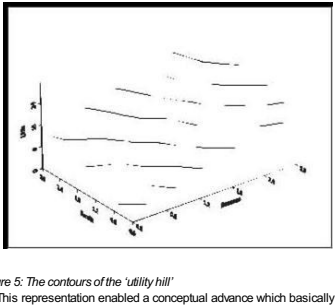


Figure 5: The contours of the 'utility hill'

This representation enabled a conceptual advance which basically gave birth to modern consumer theory – however, as we shall later, it also introduced an insurmountable intellectual dilemma. Since consumers were presumed to be motivated by the utility they gained from consumption, and points of equal utility height gave them the same satisfaction, then a consumer should be 'indifferent' between any two points on any given curve, since they both represent the same height, or degree of utility. These contours were therefore christened 'indifference curves'. This brings us to Figure 5, which is almost the way in which indifference curves are normally drawn in economic textbooks.

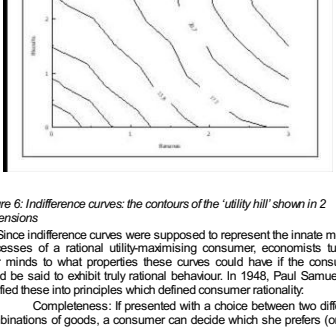


Figure 6: Indifference curves: the contours of the 'utility hill' shown in 2 dimensions

Since indifference curves were supposed to represent the innate mental processes of a rational utility-maximising consumer, economists turned their minds to what properties these curves could have if the consumer could be said to exhibit truly rational behaviour. In 1948, Paul Samuelson codified these into principles which defined consumer rationality:

w Completeness: If presented with a choice between two different combinations of goods, a consumer can decide which she prefers (or can decide that she gets the same degree of satisfaction from them, in which case she is said to be indifferent between them);

w Transitivity: If combination A is preferred to combination B, and B to C, then A is preferred to C;

w Non-satiation: More is always preferred to less. If combination A has as many of all but one commodity as B, and more of that one than B, then A is necessarily preferred to B;

w Convexity: The marginal utility a consumer gets from each commodity falls with additional units, so that indifference curves are convex in shape (shaped like a 'slippery dip').

The data in Table 2 leads to the 'indifference curves' plotted in Figure 5, which do not meet the last of these criteria, since there are 'kinks' in the contours. To qualify as rational, the consumer's indifference curves had to look like those in Figure 6; kinks are out of the question.

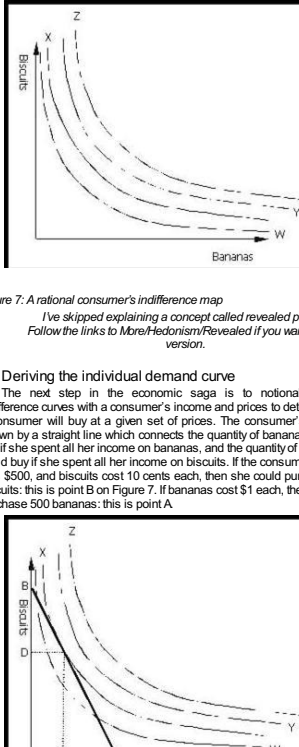


Figure 7: A rational consumer's indifference map

I've skipped explaining a concept called revealed preference. Follow the links to More/Hedonism/Revealed if you want the full version.

Deriving the individual demand curve

The next step in the economic saga is to notionally combine indifference curves with a consumer's income and prices to determine what a consumer will buy at a given set of prices. The consumer's income is shown by a straight line which connects the quantity of bananas she could buy if she spent all her income on bananas, and the quantity of biscuits she could buy if she spent all her income on biscuits. If the consumer's income was $500, and biscuits cost 10 cents each, then she could purchase 5000 biscuits: this is point B on Figure 7. If bananas cost $1 each, then she could purchase 500 bananas: this is point A.

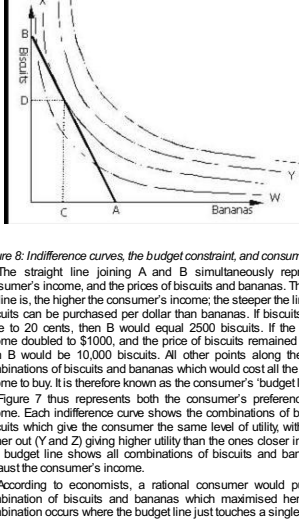


Figure 8: Indifference curves, the budget constraint, and consumption

The straight line joining A and B simultaneously represents the consumer's income, and the prices of biscuits and bananas. The further out the line is, the higher the consumer's income; the steeper the line, the more biscuits can be purchased per dollar than bananas. If biscuits doubled in price to 20 cents, then B would equal 2500 biscuits. If the consumer's income doubled to $1000, and the price of biscuits remained at 10 cents, then B would be 10,000 biscuits. All other points along the line show combinations of biscuits and bananas which would cost all the consumer's income to buy. It is therefore known as the consumer's 'budget line'.

Figure 7 thus represents both the consumer's preferences and her income. Each indifference curve shows the combinations of bananas and biscuits which give the consumer the same level of utility, with the curves further out (Y and Z) giving higher utility than the ones closer in (W and X). The budget line shows all combinations of biscuits and bananas which exhaust the consumer's income.

According to economists, a rational consumer would purchase the combination of biscuits and bananas which maximised her utility. This combination occurs where the budget line just touches a single indifference curve. If the consumer purchased any other feasible combination of biscuits and bananas using her income, then she would be foregoing some utility.[14] The point of tangency between the highest indifference curve the consumer can afford to reach – curve X in Figure 7 – and the budget line AB

therefore represents the point at which the consumer's utility is maximised, and indicates what combination of biscuits and bananas she will buy at a given set of prices for biscuits and bananas. Given the utility map, income and relative prices shown in Figure 7, the rational consumer would spend all her income to buy C bananas and D biscuits.

Have you started to fall asleep yet? Sorry, but as I warned, this stuff is boring. But have some more coffee and stay tuned; after a few more introductory bits, things start to get interesting.

The impact of changing prices on consumer demand

At this point, we have only presented the economic explanation of how a consumer will determine the consumption of any one bundle of commodities, given a fixed income and fixed prices. But what interests economists is what they call a 'demand curve', which shows how demand for a commodity changes as its price changes.

If we consider a lower price for bananas, then the number of bananas the consumer can buy rises. If at the same time her income and the price of biscuits remain constant, then the budget line moves further out on the bananas axis, but remains in the same spot on the biscuits axis. This is shown in the top half of Figure 8, and it appears as if the budget constraint 'pivots' from the point B on the biscuits axis.

Three budget lines are shown in the top half of Figure 8, each corresponding to a different price for bananas while income and the price of biscuits remain constant. Budget line I represents the highest price for bananas, since that's the price at which the consumer can buy the fewest bananas. Budget line III represents the lowest price for bananas, since that's the price at which the consumer can buy the most bananas. p1, shown on the vertical axis in the bottom half of Figure 8, is the price which generated budget line I; similarly, budget line II is associated with price p2, and budget line III with p3.

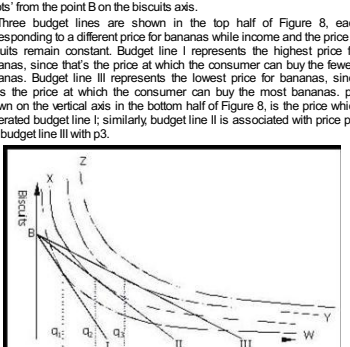


Figure 9: Deriving the demand curve

At price p1, the consumer purchases q1 bananas; at price p2, she purchases q2 bananas; and at price p3, she purchases q3 bananas. These points let us plot the individual's demand curve shown in the bottom half of Figure 8, with prices on the vertical axis and the quantities purchased on the horizontal axis.[15] We finally have a demand curve, which slopes downwards as economists predicted.[16]

I've skipped explaining a concept called compensated demand curves. Follow the links to Hedonism/Compensated if you want the full version.

The dissident Australian economist Ted Wheelwright once described this hypothesised activity as "tobogganing up and down your indifference curves until you disappear up your own abscissa", and it's easy to see why.

Nonetheless, the end result is that desired by economists: increasing a product's price will reduce a consumer's demand for that product: an individual's demand curve slopes downwards.

How rising income affects demand

So far we've looked at how economists analyse the impact of a change in prices on a consumer's behaviour. But the key flaw in the economic theory of consumer demand manifests itself in the analysis of how spending on any given commodity changes as income changes.

As with all other issues, economic theory uses indifference curves to handle this topic. The relevant commodity is placed on the horizontal axis, all other commodities on the vertical, and the budget constraint is 'moved out' (see Figure 9). This represents an increase in income with relative prices held constant – unlike a pivot, which represents a change in relative prices with income held constant. Economists say that the resulting plot – known as an 'Engels curve' – shows a consumer maximising her utility as her income rises.

One point that is essential to the approaching critique is that Engels curves can take almost any shape at all. The shapes show how demand for a given commodity changes as a function of income, and four broad classes of commodities result: necessities (Figure 9 a), which take up a diminishing share of spending as income grows; 'inferior' or 'Giffen goods', whose actual consumption declines as income rises (Figure 9 b); luxuries (Figure 9 c), whose consumption takes up an increasing share of income as it increases; and what economists term 'representative goods' (Figure 9 d), where consumption remains a constant proportion of income as income rises.

Necessities include such things as, for example, toilet paper. You will probably not buy more toilet paper – though you will certainly buy more expensive paper! – and spending on it will decline as a percentage of your income as you get wealthier.

Some products that are substitutes for better quality products when you are very poor – baked beans, perhaps – will disappear altogether from your consumption as you get wealthier, and economists refer to these as inferior goods.

Luxuries range from, for example, tourism to original works of art. Spending on holidays rises as income rises, and artworks are definitely the province of the rich.

I can't provide an example of a 'representative good', because strictly speaking, there are none. Spending on such a commodity would constitute the same percentage of income as a person rose from abject poverty to unimaginable wealth, and there is simply no commodity which occupies the same proportion of a homeless person's expenditure as it does of a billionaire's.

It may seem like explaining the obvious to say this, but the point is crucial to the approaching critique: as you age and as your income (hopefully) rises, your consumption pattern will change. Thus any 'representative consumer' is going to have a majority of curves like a and c, a smattering of curves like b, but no curves like d.

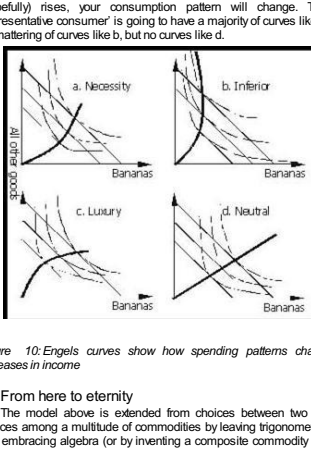


Figure 10: Engels curves show how spending patterns change with increases in income

From here to eternity

The model above is extended from choices between two goods to choices among a multitude of commodities by leaving trigonometry behind and embracing algebra (or by inventing a composite commodity called 'all other goods', as in Figure 9). Labour supply is modelled as a choice between income and leisure, with the slope of the budget line representing

the real wage. The choice between consumption now and in the future is captured by putting goods today on one axis, goods in the future on the other, and replacing relative prices with the rate of interest.[17] Indifference maps are the economist's ubiquitous, one-size-fits-all manner of representing every aspect of human behaviour.

Things fall apart

At this point, economists had gone as far as they could with the individual. To get from here to the market demand curve, they had to aggregate from a single individual to the sum of all individuals in a given economy – the aggregate formerly known as 'society'. To complete Bentham's agenda, they had to prove that a 'social indifference curve' could be constructed as a simple sum of individual utilities. The social indifference curve would accurately capture the preferences of all consumers, and enable economists to show that a market economy consisting of a multitude of self-interested, utility-maximising individuals would give rise to an outcome which maximised the welfare of all individuals.

> *Two centuries after Bentham, mathematical economists established that the second leg of the Benthamite agenda was, in general, impossible.*

The obvious conclusion from this is that Bentham was wrong: 'society' must exist as an entity in its own right, and the selfish pursuit of individual welfare does not necessarily maximise social welfare. That economists, in general, failed to draw this inference speaks volumes for the unscientific nature of economic theory.

One for all and all for one

Economists had hoped that they could prove that society acted like 'one big consumer', so that if society chose a particular bundle of commodities when given a particular set of prices, this bundle necessarily maximised 'social utility'. They hoped, in their language, that there was a social indifference map which had the same properties as the individual indifference map, and that it could be constructed simply by adding up all individual indifference maps. They hoped that market demand would obey the same laws they had derived for individuals, so that if society preferred bundle A of goods to bundle B, that choice necessarily maximised the utility of society and every individual in it. They hoped, in graphical terms, to prove the picture painted by Figure 10: that they could seamlessly move from the level of the individual to the social, and from indifference maps to demand curves.

Instead, economists proved that this consistent summation from individual to society could not be achieved. In essence, if consumers' tastes are allowed to take any form which satisfies the conditions of individual rationality, in general it will not be the case that collective behaviour is 'rational'.

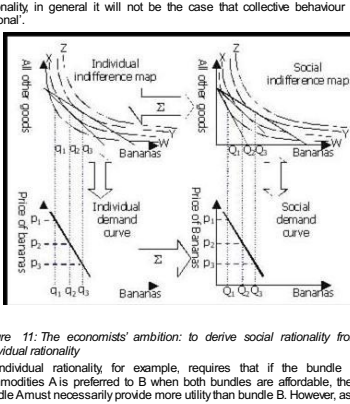


Figure 11: The economists' ambition: to derive social rationality from individual rationality

Individual rationality, for example, requires that if the bundle of commodities A is preferred to B when both bundles are affordable, then bundle A must necessarily provide more utility than bundle B. However, as I discuss below, no such result applies to a social indifference map. Economic theory therefore proves that the output level chosen by a market economy may not, and probably does not, maximise social welfare demand curves.

Nor does a market demand curve derived by summing many individual behave identically to a single individual's demand curve. An individual demand curve slopes downwards; the same cannot be said for a market

demand curve – it is quite possible for this curve to be flat in places, or even upwards sloping. This failure is put fairly bluntly in Varian's Microeconomic Analysis, which almost all academic economists have read in their Masters or PhD programs:

> Unfortunately . . . The aggregate demand function will in general possess no interesting properties . . . The neoclassical theory of the consumer places no restrictions on aggregate behaviour in general. (Varian 1992)

It took advanced mathematical logic to reach this conclusion, but the essence of the argument is easy to visualise by working backwards from what economists had hoped to prove.

You will recall that any one individual's indifference curves are the contours on a map of the utility 'hill' that the individual climbs by consuming commodities (Figures 3 and 4). Since utility is subjective, the actual height of the hill can't be specified because utility can't be quantified – there is no objective measure. However, since economic theory ignores the question of where consumer tastes come from, and assumes that income (the 'budget line') and tastes (the 'indifference curves') are independent, each individual's utility hill can be treated as constant. The same can't be said for the 'social utility hill', since if income is redistributed between individuals, then there is no way to say whether this results in a greater or lesser degree of total utility. The redistribution of income might effectively transfer one banana from someone who gets very little utility from it to someone who gets a great deal of utility – since what gives great utility to one individual may give very little to another.

There are therefore many 'social utility hills', one for every distribution of income. There are two reasons for this. Firstly, changing the distribution of income between individuals changes the weighting of each individual in the aggregate called society. Secondly, any one individual's purchases will alter as her income grows – remember that Engels curves, which plot spending as a function of income, can take almost any shape (if you are poor and therefore spend a lot of your money on instant coffee, then a change in its price has a large impact on your welfare; but if you are rich and only buy ground coffee, then a change in the price of instant coffee will have no impact on your welfare).

This issue wouldn't be such a problem if the distribution of income was fixed, but one essential aspect of conventional economic theory is that the price system determines the distribution of income. Altering prices therefore alters the distribution of income – but to be able to construct a market demand curve, you have to alter prices. There will therefore be a different social indifference map for every different set of prices. The 'social budget line' and the social indifference surface are therefore interdependent, since every set of prices will generate a different social utility hill.

To give a visual analogy, ignore the problem that there is no objective measure of utility like 'utils', and imagine that you are standing on the social utility hill that applies when a banana costs five times as much as a biscuit, at the location '1 million biscuits, 100,000 bananas'. If the relative price changes to 5.1 biscuits to the banana, the hill you were standing on is abruptly replaced by another, which could be higher or lower than its predecessor – you could suddenly 'plunge' thousands of feet, or find yourself entombed deep underground.

This is represented in Figure 11. The continuous line indifference curves and budget line indicate the contours of the social utility hill at the first price ratio aa; the dotted line indifference curves and budget line indicate its contours at the new price ratio ab. These two sets of indifference curves contours 'intersect', which is by definition impossible for an individual, but quite possible when you try to aggregate the behaviour of two or more individuals. If you tried to generate an overall social utility map by considering all possible prices for bananas and biscuits, it would therefore be extremely jagged. Rather than being the smoothly upward sloping hill that economists visualise, it would look more like a mountain range undergoing a catastrophic earthquake. The resulting set of indifference curves, rather than sloping smoothly, might resemble numerous contour maps of the Himalayas from different geological periods, superimposed haphazardly over each other.

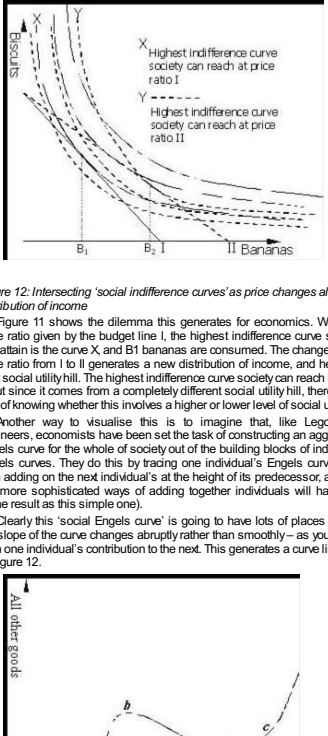


Figure 12: Intersecting 'social indifference curves' as price changes alter the distribution of income

Figure 11 shows the dilemma this generates for economics. With the price ratio given by the budget line I, the highest indifference curve society can attain is the curve X, and B1 bananas are consumed. The change in the price ratio from I to II generates a new distribution of income, and hence a new social utility hill. The highest indifference curve society can reach is now Y, but since it comes from a completely different social utility hill, there's no way of knowing whether this involves a higher or lower level of social utility.

Another way to visualise this is to imagine that, like Lego-block engineers, economists have been set the task of constructing an aggregate Engels curve for the whole of society out of the building blocks of individual Engels curves. They do this by tracing one individual's Engels curve, and then adding on the next individual's at the height of its predecessor, and so on (more sophisticated ways of adding together individuals will have the same result as this simple one).

Clearly this 'social Engels curve' is going to have lots of places where the slope of the curve changes abruptly rather than smoothly – as you move from one individual's contribution to the next. This generates a curve like that in Figure 12.

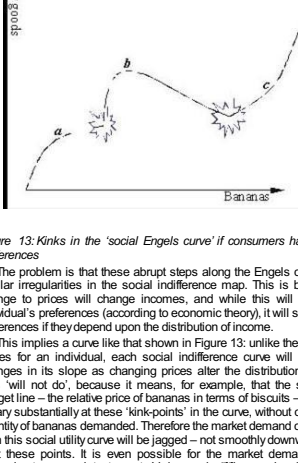


Figure 13: Kinks in the 'social Engels curve' if consumers have different preferences

The problem is that these abrupt steps along the Engels curve require similar irregularities in the social indifference map. This is because any change to prices will change incomes, and while this will not shift an individual's preferences (according to economic theory), it will shift society's preferences if they depend upon the distribution of income.

This implies a curve like that shown in Figure 13: unlike the indifference curves for an individual, each social indifference curve will have abrupt changes in its slope as changing prices alter the distribution of income. This 'will not do', because it means, for example, that the slope of the budget line – the relative price of bananas in terms of biscuits – will be able to vary substantially at these 'kink-points' in the curve, without changing the quantity of bananas demanded. Therefore the market demand curve derived from this social utility curve will be jagged – not smoothly downward sloping – at these points. It is even possible for the market demand to slope upwards at some points: to quote Varian again, "[t]he neoclassical theory of the consumer places no restrictions on aggregate behaviour in general".

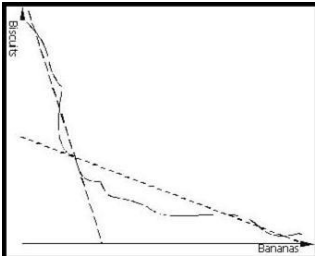


Figure 14: Discontinuities in a 'social utility curve' constructed by aggregating valid individual indifference curves: no change in demand for a change in prices

Far from proving Bentham's supposition that "the interests of the community then is (sic)the sum of the interests of the several members who compose it", economists instead showed that only under highly restrictive conditions could social welfare be treated as the sum of its individual members. These were that:

w The distribution of income was fixed and Engels curves must have a constant slope; or

w Engels curves must have a constant slope, and they all have the same slope.

The first restriction was unacceptable because this contradicts the economic theory of income distribution, which argues that relative incomes are determined by the price system. So the only 'palatable' way out for economists was with the second restriction.

However, the economic meaning of the first half of that restriction is that all Engels curves must be straight lines, like the Engels curves in diagram d of Figure 9, so that all commodities are neither luxuries nor necessities (Engels 'curves' for any single consumer can be of any length, so that incomes vary between individuals, but all 'curves' must be straight lines). But we earlier concluded that there would be few, if any, goods in this category. Yet economists have to pretend that increasing the income of a single consumer by, say, a factor of 100, increases her consumption of all commodities by the same factor.

The economic meaning of the second half of the restriction is, if anything, even more extreme than that of the first half. It is that every consumer has to spend the same proportion of each new dollar of income the same way, which means that all consumers must have the same tastes. 'We are all individuals', and, like the crowd in Monty Python's Life of Brian, 'we are all the same'.

Individual indifference curves therefore cannot take any allowable shape and still enable economists to treat society as the sum of the individuals in it. Instead, they must all take shapes which generate Engels curves like those shown by the lines marked I, II and III in Figure 14. These straight-line Engels 'curves' meet the restrictions imposed in condition 2 (in the jargon, they are known as 'homothetic' and 'affine').

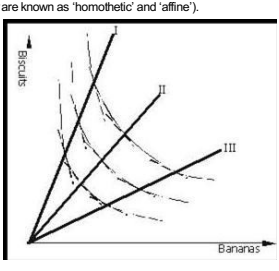


Figure 15: The only allowable utility curves if individual utilities are to be summed to yield social utility

Think what these two additional restrictions mean in practice. They

mean, for example, that Bill Gates must spend the same proportion of every new dollar he earns in exactly the same way that you do; that you must allocate each new dollar you earn in the same way that Bill Gates does.

Let's imagine that Bill, in his youth, was a typical computer programmer – in that he virtually lived on pizza: let's say he spent 10% of any increase in his income on pizzas. If his income went up by $1000 in a year, he'd spend $100 of that change on pizzas.

Imagine how many pizzas Bill Gates must have consumed since 1990 if this were true.

Working in the reverse direction, try to calculate how many fragments of Leonardo da Vinci's Codex Leicester you would have in your possession, if you had spent as much of your additional income since 1990 as Gates has on acquiring this antiquity. Whichever way you look at it, this assumption – that all consumers have the same tastes, and that consumption patterns don't change with income – is absurd.

It means that as you get older, you continue spending your income exactly as you did when you were a teenager. If you once spent 10% of each new dollar income on cola, and you earned $1000 a year two decades ago and $1,001,000 now, then today your annual expenditure on cola must be at least $100,000. Similarly, if you once spent nothing on health care, then you will never spend anything on health care.

The absurdity of these conditions is obvious. One of the many ways we are distinguished from each other is how our consumption preferences differ; to assume that we have exactly the same preferences is to assume that there is only one individual in society. One of the many ways in which we are human is that we change, and the consumptions of youth give way to different expenditures as we grow older.[18] To assume that we never change is to ignore our humanity.

> *The economic theory of consumer behaviour collapses then to this: if the market economy has only one consumer, and that consumer only ever consumes the one commodity, then individual utility can be summed to yield social utility.*

Any sane person would at this point abandon the initial theory that society can be treated as a simple sum of the individuals in it (I discuss what this sane response means in Chapter 12). Instead, economists actually accept the restrictions cited above: that either all individuals are identical, or society consists of just one individual!

Varian's behaviour on this issue is unfortunately typical of economists. Though he initially informs budding academic economists that economic theory cannot derive a downward-sloping demand curve from its theory of individual behaviour, he suggests that the way out of the dilemma is to:

> Suppose that all individual consumers' indirect utility functions take the Gorman form . . . [where] . . . the marginal propensity to consume good j is independent of the level of income of any consumer and also constant across consumers . . . This demand function can in fact be rationalized by a representative consumer . . . (Varian 1984; emphasis added. The 3rd edition, published in 1992, replaces 'rationalized' with the more reasonable-sounding 'generated'.)

This rationalisation of the 'representative consumer' permeates modern economics – it has even taken over macroeconomic analysis, so that economists model an entire economy as if there is only one person in it (which they describe by the more general term of 'representative agent'). Many academic economists doubtless believe that the representative agent has been shown to be a valid abstraction. Yet far from being valid, it is in fact a fudge devised to get around the failure to prove that society can be reduced to the sum of its constituent individuals.

Divisions in the flock

In an example of the gulf that exists between leading mathematical economists and their lesser brethren, the researchers who discovered this failure are not nearly as sanguine as Varian about the rationalisation used to sidestep it.[19]

> First, when preferences are homothetic and the distribution of income (value of wealth) is independent of prices, then the market demand function (market excess demand function) has all the properties of a consumer demand function . . . Second, with general (in particular non-homothetic) preferences, even if the distribution of income is fixed, market demand functions need not satisfy in any way the classical restrictions which characterize consumer demand functions . . . The importance of the above results is clear: strong restrictions are needed in order to justify the hypothesis that a market demand function has the characteristics of a consumer demand function. Only in special cases can an economy be expected to act as an 'idealized consumer'. The utility hypothesis tells us nothing about market demand unless it is augmented by additional requirements. (Shafer & Sonnenschein 1982)

These 'strong restrictions' are that society is populated by a horde of clones, consuming commodities which are identical to each other.

Both assumptions are patent nonsense. The only sensible reaction to these theorems is to accept that the entire Benthamite endeavour is invalid: society is something more than the sum of its individual members. When this was established, economists should have downed tools and started designing all over again, from a fundamentally different foundation. But they did not.

Instead, many of those who discovered this failure allowed themselves to accept absurd assumptions about individual and collective behaviour. These assumptions are stated with sometimes breathtaking nonchalance. Consider the following examples:

> The necessary and sufficient condition quoted above is intuitively reasonable. It says, in effect, that an extra unit of purchasing power should be spent in the same way no matter to whom it is given. (Gorman 1953)

> Two criteria will be considered which lead to the possibility of aggregation: (1) identical preferences (hence identical demand functions), and (2) proportional incomes . . . These results have a number of interesting applications in the pure theory of international trade. (Chipman 1974)

These assumptions have since become part of the standard fare of advanced economic education (where they are known as the Sonnenshein-Mantel-Debreu or SMD conditions), and are normally taught without even discussing how manifestly unrealistic they are.[20]

Following the madding crowd

There are many reasons why economists did not recoil from the patent absurdities outlined above, and search for a sounder approach to economic theory than Bentham's individualistic calculus.

One is that economics has been wedded to the vision of society as simply a sum of utility-maximising individuals since the inception of neoclassical economics in the 1870s. When the proof came, one century later, that this vision was internally inconsistent, the commitment to the vision was too strong to break. Better to search for special conditions which could let the theory survive – however ludicrous they might be – than to admit failure.

A second reason is that the peculiar language and mathematics used to derive these results makes it difficult to see just how absurd the assumptions needed to sustain the aggregation process are. It sounds much more highbrow to say that 'preferences are assumed to be homothetic and affine in income' than it does to say 'we assume all consumers are identical and never change their spending habits'.

A third reason, perhaps the key one, is the division of mainstream economists into effective 'castes', with only a tiny but exalted subset of the profession undertaking the detailed mathematical work needed to discover the weaknesses in the theory. The vast majority of economists believe that this high caste, the mathematical economists, did their work properly, and proved that the theory is internally consistent. The caste has indeed done its work properly, but it has proven precisely the opposite: that the theory is only consistent under the most restrictive and specious of assumptions.

However, rather than taking the next logical step, and acknowledging that the foundations of economics are unsound and must therefore be changed, most mathematical economists are so wedded to this way of thinking, and so ignorant of the real world, that they instead invent some fudge to disguise the gaping hole they have uncovered in the theory.

The majority of economists, blithely unaware of this state of affairs, then accept this fudge by the Brahmins of the profession as faithfully as Hindus accept the cleansing properties of the Ganges River. As a result, the fudge then turns up in more mundane areas of economics, such as 'macroeconomics' (discussed in Chapter 9), where economists today analyse the economy as if it consisted solely of a single representative agent.

Consequently, these supposedly more practical theories can provide zip guidance in the serious business of managing a market economy. You would do as well to consult a ouija board as an economist who rigorously follows economic theory when giving advice.

The Sonnenshein-Mantel-Debreu result is the first of many that we will consider which have effectively split the caste of mathematical economists into two sects. One pretends that business as usual can continue, despite the presence of this (and many other) fallacies in the creed. The other is dabbling in alternative religions – such as complexity theory or evolutionary economics.

Sadly, the uninformed majority of the profession believes that the first sect is the bearer of the true religion, and that the members of the second sect have betrayed the faith. A more accurate analogy is that the dabblers in alternative religions are experiencing the first flushes of adolescence, while

the majority of the profession remains mired in infancy. Clearly, the Benthamite ambition to portray society as simply an aggregate of its individual members is a failure. The whole is more than the sum of the parts.

The neoclassical rejoinder

The great irony of this particular critique of economics is that it was constructed by its supporters. There is, as a result, no articulate rejoinder. Instead there are rationalisations, such as the 'representative agent' – which, as in Varian 1984, are often openly described as such.

If a defence were to be given of this practice, it would probably be what Samuelson termed 'the ': that the assumptions of a theory don't matter, instead all that counts is how accurately a theory predicts reality. This popular but clearly invalid methodological defence is debunked in Chapter 7.

So What?

It might seem strange to make such a song and dance about whether market demand curves slope downwards. While economic theory clearly fails to prove that market demand falls smoothly as price rises, there are some sound reasons why demand might generally be a negative function of price. For example, a rise in the price of a commodity can force poorer consumers to substitute some cheaper alternative – or go without.[21] So why does it matter that economists can't prove this?

Firstly, it matters because economists are trying to prove that a market economy necessarily maximises social welfare. If they can't prove that the market demand curve falls smoothly as price rises, they can't prove that the market maximises social welfare.

Secondly, the concept of a social indifference curve is crucial to many of the key notions of economics: the argument that free trade is necessarily superior to regulated trade, for example, is first constructed using a social indifference curve. Therefore, if the concept of a social indifference curve itself is invalid, then so too are many of the most treasured notions of economics. As even Gorman put it:

> Community indifference and utility possibility loci are among the most useful concepts of welfare economics. Their great disadvantage is that they may intersect . . . Thus the analysis . . . frequently becomes inconclusive. (Gorman 1953)

This 'disadvantage' formally invalidates much of economic theory, and with it policy recommendations based upon that theory.

This is because if we take the Sonnenshein-Mantel-Debreu conditions seriously, there must be discontinuities in any social indifference map in any society with more than one individual. Market demand curves should therefore be similarly jagged. Rather than the definitive downward-sloping curve of the standard totem, economic theory cannot rule out a shape like Figure 15.

Aesthetics aside, one of the many problems which such a curve presents for economic theory is that it can intersect the supply curve in more than one place, as shown. This possibility undermines one of the key articles of the economic faith, that 'everything happens in equilibrium'. If there are multiple points of intersection between supply and demand, there will be multiple points where 'everythifng happens'. How then can you determine which will prevail in practice, let alone decide whether any one equilibrium is better or worse than any other?

These dilemmas flow from what appeared at the time to be a conceptual advance – dropping the fiction that utility could be measured in units akin to those we use to gauge weight, etc. While this was indeed more realistic, its interaction with two other aspects of economic theory made it impossible to aggregate the utility of two or more individuals.

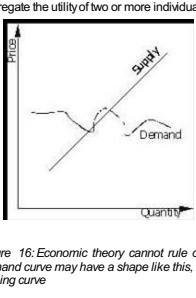


Figure 16: Economic theory cannot rule out the possibility that a market demand curve may have a shape like this, rather than a smooth, downward-sloping curve

The culprits are the highly subjective nature of the concept of utility, and the belief that the price system determines income distribution. Since a change in relative prices will change the distribution of income, it therefore changes who consumes what, and hence the 'sum' of the subjective utility of all individuals. Since utility is subjective,[22] there is no way to determine whether one distribution of income generates more or less aggregate utility than any other.

Economists originally used this aspect of their theory to argue against social reformers who wished to redistribute income from the rich to the poor. They argued that such a redistribution might actually reduce social welfare by taking a unit of a commodity from a rich person who derived a great deal of utility out of it, and giving it to a poor person who derived very little utility from it.

It is ironic that this ancient defence of inequality ultimately backfires on economics, by making it impossible to construct a market demand curve which is independent of the distribution of income. If the market demand curve depends upon the distribution of income, if a change in prices will alter the distribution of income, and if this does not result in a single equilibrium between demand and supply, then economics cannot defend any one distribution of income over any other. A redistribution of income that favours the poor over the rich cannot be formally opposed by economic theory.

Finally, this failure rehabilitates the approach of classical economics to analysing the behaviour of individuals in society. Classical economists such as SmithSmith, Ricardo and Marx divided society into social classes, and considered how different policies might favour one social class over another. The notion of class has been expunged from economics by the concept of the indifference curve and its 'one size fits all' treatment of everyone from the poorest Bedouin to the richest American. Yet because the preferences of different individuals cannot be meaningfully aggregated, this concept is invalid for the analysis of anything more than an isolated individual.

But the conditions under which aggregation is valid – when tastes are identical and unaffected by changes in income – are at least reasonable as first approximations when the analysis splits society into different social classes. It is not too unreasonable to lump all workers, all landlords, and all capitalists together, as Smith, Ricardo and Marx used to do. Incomes within a class vary substantially less than incomes between classes, and tastes are far more likely to be common within classes than between them. A model with both Robinson Crusoe and Friday is at least slightly more reasonable than a model with Robinson Crusoe alone.

Leading mathematical economists have made very similar musings to this. Alan Kirman made one of the strongest such statements in his provocatively titled paper 'The intrinsic limits of modern economic theory: the emperor has no clothes' (Kirman 1989).[23] After discussing these and other theoretical failures of neoclassical economics, Kirman concluded that

> If we are to progress further we may well be forced to theorise in terms of groups who have collectively coherent behaviour. Thus demand and expenditure functions if they are to be set against reality must be defined at some reasonably high level of aggregation. The idea that we should start at the level of the isolated individual is one which we may well have to abandon. (Kirman 1989)

But Wait, There's More

With the conclusions above evident in the mainstream economic literature, one hardly needs attacks by critics to demolish the notion of utility maximisation. Nonetheless, of course, criticisms abound. These include:

w The non-evolutionary way in which consumer behaviour is analysed, when the important issues in consumption concern the way tastes and the system of production co-evolve over time;

w The schizophrenic vision of how consumers behave – maximising their utility on the one hand, while scrupulously adhering to ethical contractual behaviour on the other;

w The irrationality of the economic definition of rational, which assumes we have infinite processing power to calculate the comparative utilities of different bundles of goods. In fact, the rational thing to do when faced with overwhelming choice is to minimise choice by relying upon habit;

w The misplaced focus upon individuals as self-oriented utility maximisers may ignore the most essential difference between humans and other species – the development of ethical behaviour.

These issues are explored in detail on the web. Follow the link to More/Hedonism.

The failure of the economic theory of demand to achieve its objective – to represent society as no more than the sum of its individuals – also means that the economic picture of an individual as entirely self-interested is both fallacious and superfluous. Indeed, real progress in understanding consumption may require economists to explicitly analyse the ways in which the behaviour and welfare of other people affects individual behaviour. This

is being done by some economists, in an area now described as 'network economics' – because an individual's 'network' of relationships affects her purchases (see, for example, Oda et al 2000, and Yao and Darwen 2000).

However, most economists continue to behave as if their Benthamite perspective on individuals and consumption is justified – because their selective and disjointed education either omits discussion of the fatal flaws in the Benthamite vision or trivialises the true meaning of these logical conundrums. In general, economists are equally confident of the other half of the pair of scissors, the supply curve. However, the next chapter shows that this other blade is almost as fragile as its demand sibling.

3. The price of everything and the value of nothing

Why most products cost less to produce as output rises

We have already seen that the demand half of conventional economic analysis is unsound: economics cannot derive a smooth, downward-sloping market demand curve from its theory of individual behaviour. But on the supply side, surely it makes sense that, to elicit a larger supply of a commodity, a higher price must be offered?

There is, in fact, an alternative proposition, which held sway in economics for its first century. This was the argument of the classical school of economics that price was set by the cost of production.[24] When this proposition is put in the same static form as economics uses to describe a commodity market, it translates as a horizontal (or even a falling) supply curve, so that the market price doesn't change as the quantity produced rises (and it can actually fall). This chapter shows that, though the modern position is superficially more appealing and apparently more sophisticated, there are logical problems with it which mean that the classical position is a more accurate description of reality.

The kernel

One of the peculiar aspects of modern Western society is that the majority of the population has no direct experience of how the commodities it consumes are produced. Only a small and decreasing minority is directly involved in production, and only a minority of that minority has direct knowledge of how factories are designed and managed. In contrast to consumption, the conditions under which commodities are produced are a mystery to most people, and the economic analysis of production appears to illuminate that mystery.

Economic theory argues that productivity falls as output rises, so that higher levels of output result in higher prices. The 'supply curve' therefore slopes upwards: a higher price has to be offered to entice firms to produce a higher output.

Though this sounds intuitively plausible, when this theory was put to those who do know how factories are designed and managed, they rejected it as "the product of the itching imaginations of uninformed and inexperienced arm-chair theorizers" (Lee 1998, citing Tucker).

How could something which seems so reasonable to the inexperienced, be so absurd according to those 'in the know'? The answer lies in the assumptions economists make about production. Though these seem sound to the uninitiated, the two key assumptions are in fact contradictory: if one applies for a given industry, then the other one almost certainly does not.

Economic theory also doesn't apply in the 'real world' because engineers purposely design factories to avoid the problems that economists believe force production costs to rise. Factories are built with significant excess capacity, and are also designed to work at high efficiency right from low to full capacity. Only products (like oil) that can't be produced in factories or farms are likely to have costs of production that behave the way economists expect.

The outcome is that costs of production are normally either constant or falling for the vast majority of manufactured goods, so that 'supply curves' are normally either flat or downward sloping. This causes manufacturers no difficulties, but it makes life impossible for economists, since most of economic theory depends on supply curves sloping upwards.

The roadmap

In this chapter I outline the economic analysis of production, which concludes that productivity will fall as output rises, leading to rising costs of production. This in turn leads to a need for the market price to rise if producers are to supply more, which economists represent as a 'supply curve' that slopes upward in price.

Next I detail Sraffa's argument that two crucial assumptions of this analysis – that supply and demand are independent, and that at least one input to production can't be varied in the short run – are mutually exclusive. A number of potential economic rejoinders to this argument are considered and dismissed. The economic argument that profit is maximised by equating 'marginal cost' to 'marginal revenue' is shown to be incorrect once time is taken into account.

The outline of the economic analysis of production below will probably convince you that the theory makes sense, but as with the corresponding section in Chapter 2, it is almost certain to bore you senseless. It is also unavoidably laden with jargon, and less accessible than Chapter 2 since few of us have any experience of production to the same depth as we have experience of consumption. So go back to the coffee pot, pour yourself a strong one, and read on.

Diminishing productivity causes rising price

The economic theory of production argues that capacity constraints play the key role in determining prices, with the cost of production – and therefore prices – rising as producers try to squeeze more and more output out of a fixed number of machines. The argument has several stages: stage one puts the proposition that productivity falls as output rises; stage two takes the declining productivity argument and rephrases it as rising costs; and stage three determines the point of maximum profitability by identifying where the gap between revenue and costs is greatest.

Stage one: productivity falls as output rises

Economic theory asserts that the supply curve slopes upward because productivity falls as output rises. This falling productivity translates into a rising price. There is thus a direct link between what economists call 'marginal productivity' – the amount produced by the last worker – and 'marginal cost' – the cost of producing the last unit.

Table 1 shows an example of how economists conceive of production. This mythical firm has fixed costs of $250,000, and pays its workers a wage of $1000.[25] It can sell as many units as it can produce at the market price of $4. To produce output at all, the firm must hire workers: with no workers, output is zero. The first worker enables the firm to produce 52 units of output. This is shown in the first row of the table: the labour input is one unit, and total output is 52 units.

The marginal product of this worker – the difference between production without him/her (zero) and production with – is 52 units. The marginal cost of the output is the worker's wage – $1000 – divided by the number of units produced – 52 – which yields a marginal cost of $19.20.

The average fixed costs of output at this point are enormous – $250,000 divided by just 52, or $4807. The average total cost is $251,000 divided by 52, or $4827 per unit – which implies a loss of $4823 per unit sold, if this were the chosen level of production.

Labour	Output	Wage Bill	Total Cost	Marginal Product	Marginal Cost	A Variab Cost
1	52	1000	251000	52.0	19.2	1
9	611	9000	259000	83.6	12.0	1
10	698	10000	260000	87.5	11.4	1
100	23333	100000	350000	398.5	2.5	4
276	131111	276000	526000	772.5	1.3	2
277	131885	277000	527000	773.7	1.3	2
400	233333	400000	650000	850.0	1.2	2
500	316667	500000	750000	800.5	1.2	2
700	443333	700000	950000	401.5	2.5	2
725	452370	725000	975000	323.5	3.1	2
730	453938	730000	980000	307.1	3.3	2
735	455424	735000	985000	290.5	3.4	2
740	456827	740000	990000	273.7	3.7	2
745	458144	745000	995000	256.6	3.9	2
746	458397	746000	996000	253.1	4.0	2
747	458647	747000	997000	249.7	4.0	2
748	458893	748000	998000	246.2	4.1	2
800	466667	800000	1050000	52.0	19.2	2

Table 1: Input and output data for a hypothetical firm

At this stage, production benefits from economies of scale. Just one worker had to perform all tasks, whereas a second worker allows them to divide up the jobs between them, so that each specialises to at least that extent. With specialisation, the productivity of both workers rises. The same process continues with the 9th and 10th workers, so that the marginal product of the 9th – the amount he adds to output over and above the amount produced by 8 workers – is 83.6 units. Similarly, the marginal product of the 10th worker is 87.5 units.

If the firm actually produced this number of units, it would lose $257,207 dollars – more than its fixed costs. However, the process of rising marginal productivity – and therefore falling marginal cost – comes to the rescue as output rises. By the 100th worker, the firm is still making a loss, but the loss is falling because its marginal cost has fallen below the sale price. The 100th worker adds 398.5 units to output, at a marginal cost of $1000 divided by 398.5, or just $2.50 a unit. This is less than the sale price of $4 a unit, so the firm is making a profit on the increase in output – but only enough to reduce its losses at this stage, rather than to put it into the black.

Black ink arrives with the 277th worker, who brings in $3090 profit – the

proceeds of selling the 772.5 additional units the worker produces at the sale price of $4 a unit – for the cost of his wage of $1000.

This process of rising marginal productivity continues right up to the 400th worker hired. By this stage, marginal cost has fallen dramatically. The 400th worker adds 850 units to output, so that the marginal cost of his output is the wage of $1000 divided by 850, or $1.18 (rounded up to $1.2 in the table). Average fixed costs, which were enormous at a tiny level of output, are relatively trivial at the output level of 233,333 units: they are down to just over a dollar.

From this point on, productivity of each new worker ceases to rise. Each new worker adds less to output than his predecessor. The rationale for this is that the ratio of workers – the 'variable factor of production' – to machinery – the 'fixed factor of production' – has exceeded some optimal level. Now each extra worker still adds output, but a diminishing rate. In economic parlance, we have reached the region where diminishing marginal productivity applies. Since marginal product is now falling, marginal cost will start to rise.

But profit continues to rise, because though each additional worker adds less output and therefore brings in less revenue, the revenue from the additional units still exceeds the cost of hiring the worker. In economic parlance, marginal revenue exceeds marginal cost.

We can see this with the 500th worker, who adds 800.5 units to output. The marginal cost of his output is the wage ($1000) divided by 800.5, or $1.25 (rounded down in the table). This is higher than the minimum level of $1.18 reached with the 500th worker. But the additional units this worker produces can all be sold at $4, so the firm still makes a profit out of employing the 500th worker.

The same principle still applies for the 600th worker, and the 700th. Productivity has dropped sharply now, so that this worker adds only 401.5 units to output, for a marginal cost of $2.50. But this is still less than the amount the additional output can be sold for, so the firm makes a profit out of this worker.

This process of rising profit comes to an end with the 747th worker, whose additional product – 249.7 units – can only be sold for $998.8, versus the cost of his wage of $1000. From this point on, any additional workers cost more to employ than the amount additional output they produce can be sold for.

The firm should therefore employ 746 workers, and maximise its profit at $837,588. At this point, the marginal cost of production equals the marginal revenue from sale, and profit is maximised.

The 800th adds 52 units, for a now soaring marginal cost of $19.20. By the time we get to the 812th worker, workers are – metaphorically speaking – falling over each other on the factory floor, and this workers adds a mere 3.3 units to output, for a marginal cost of $300. The next worker actually reduces output.

The next few figures put this table into graphical form. Figure 1 puts labour inputs on the horizontal axis, output on the left-hand vertical axis (with a scale from zero to 500,000) and marginal product on the right hand axis (with a scale from zero to 5000 – but only the bottom range, to 1000, is relevant). As also shown in the fifth column of Table 1, marginal product – the output generated by each new worker – rises for quite a while. It reaches its highest level at the 400th worker, who adds 850 units to output (shown by the dotted line running to the right hand vertical axis from where A's arrow cuts the lower of the two curves). From that point on, the slope of the output curve starts to fall. By the time we have reached the 813th worker, we reach the maximum possible level of output, point B. Here the marginal product of the last worker is almost zero, and the marginal product of the next worker is actually negative.

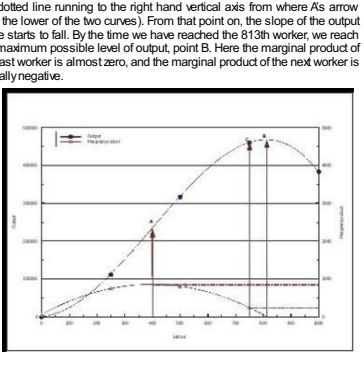


Figure 1: Product per additional worker falls as the number of workers hired

rises

Stage two: falling productivity means rising cost

Economists normally rush straight into this argument in their textbooks, but I will make the transition from stage one to stage two slowly, to emphasise the point that to economists, everything is determined by diminishing marginal productivity.

The first step in this process is Figure 2, which simply swaps the axes around from Figure 1. Whereas the labour input was on the horizontal axis and output on the vertical in Figure 1, output in Figure 2 is on the horizontal axis and labour input on the vertical. Pick any level of output on the horizontal axis of Figure 2, and the vertical axis tells us how many workers are needed to produce it.

Next, we convert the vertical axis in Figure 2 from a measure of labour to a measure of variable costs. To do this, all we have to do is multiply the labour input by the wage. Since in this example the wage is $1000, all the numbers on the vertical axis in Figure 3 are 1,000 times the numbers in Figure 2.

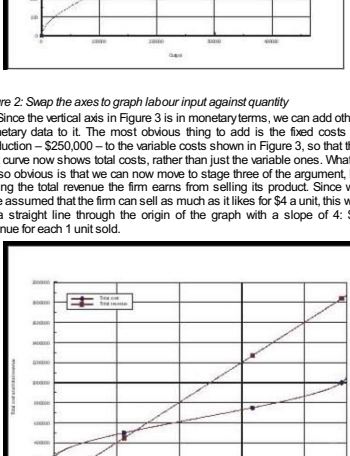


Figure 2: Swap the axes to graph labour input against quantity

Since the vertical axis in Figure 3 is in monetary terms, we can add other monetary data to it. The most obvious thing to add is the fixed costs of production – $250,000 – to the variable costs shown in Figure 3, so that the cost curve now shows total costs, rather than just the variable ones. What's not so obvious is that we can now move to stage three of the argument, by adding the total revenue the firm earns from selling its product. Since we have assumed that the firm can sell as much as it likes for $4 a unit, this will be a straight line through the origin of the graph with a slope of 4: $4 revenue for each 1 unit sold.

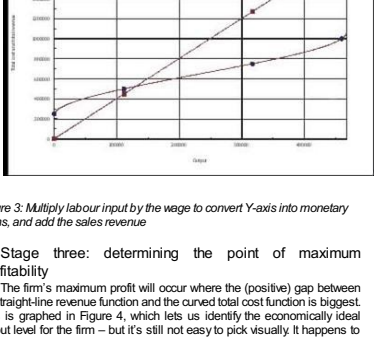


Figure 3: Multiply labour input by the wage to convert Y-axis into monetary terms, and add the sales revenue

Stage three: determining the point of maximum profitability

The firm's maximum profit will occur where the (positive) gap between its straight-line revenue function and the curved total cost function is biggest. This is graphed in Figure 4, which lets us identify the economically ideal output level for the firm – but it's still not easy to pick visually. It happens to

be where the slope of the total revenue curve equals the slope of the total cost curve, because this results in the biggest gap between total revenue and total cost.

This is because when the revenue curve rises more steeply than the cost curve, an additional sale will make the gap bigger still. When the cost curve rises more steeply than the revenue curve, an additional sale will reduce the gap between revenue and costs. Therefore the biggest gap occurs where the slopes of the two curves are equal. This is shown in the plot of total profit in Figure 4. If you check Table 1, you can see that the profit column reaches its maximum of $837,588 at 746 workers. The same maximum is just apparent in Figure 4 at the far right-hand end of the profit curve (and the graph turns down very rapidly after that).

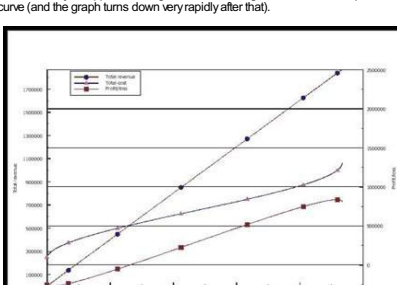


Figure 4: Maximum profit occurs where the gap between total cost and total revenue is at a maximum

While this graph does tell us the level of output at which profit is maximised, economists prefer to locate the point of maximum profit by working in terms of the slopes of the cost and revenue curves – marginal cost and marginal revenue – rather than the curves themselves.

The slope of the total revenue curve is easy to work out in this example: it's simply the price of $4 per unit sold (a monopoly is more complicated, but we'll leave that problem to Chapter 4). The slope of the total cost curve can't be worked out easily from any earlier figure, since it is equal to the change in cost divided by the change in output. But we did work this out in Table 1, so those numbers are used to produce the next two figures. Total cost and marginal cost are shown in Figure 5, with the data points A, B and C shown once again for reference. Remember that point A, with 400 workers employed, gave the highest marginal product and therefore the lowest marginal cost – you can see this from the right-hand axis of the figure. You can also see that point C, with 746 workers employed, has a marginal cost of $4 – which is equal to the sale price.

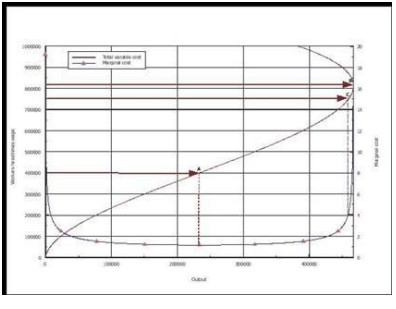


Figure 5: Deriving marginal cost from total cost

Finally, we can plot the standard diagram economists use to represent the behaviour of a firm in Figure 6. As well as plotting marginal cost and marginal revenue, Figure 6 also shows: average fixed cost, which starts off

astronomically high, and falls smoothly towards zero as output rises; average variable cost, which falls as productivity rises initially, but then starts to rise once marginal cost exceeds average cost; and average total cost, which combines average fixed and variable costs. The slope of the revenue curve is called marginal revenue, while the slope of the cost curve is called marginal cost.

The point where the two marginal curves intersect corresponds to the point at which the distance between the total revenue and total cost curves is greatest. We have arrived at the economic Nirvana: a perfectly competitive firm (the concept of 'perfect competition' is discussed in detail Chapter 4) producing a quantity at which marginal cost equals both marginal revenue and price.

A different diagram

Students of economics will note that Figure 6 is rather different from the normal diagram drawn in economics textbooks: the marginal cost curve rises much less steeply in the standard textbook graphs than it does here, while average cost rises much more steeply in textbook drawings, once marginal cost exceeds it.

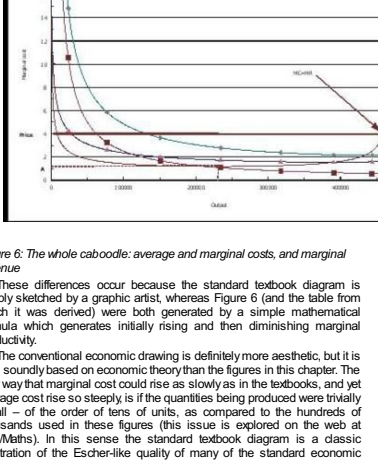


Figure 6: The whole caboodle: average and marginal costs, and marginal revenue

These differences occur because the standard textbook diagram is simply sketched by a graphic artist, whereas Figure 6 (and the table from which it was derived) were both generated by a simple mathematical formula which generates initially rising and then diminishing marginal productivity.

The conventional economic drawing is definitely more aesthetic, but it is less soundly based on economic theory than the figures in this chapter. The only way that marginal cost could rise as slowly as in the textbooks, and yet average cost rise so steeply, is if the quantities being produced were trivially small – of the order of tens of units, as compared to the hundreds of thousands used in these figures (this issue is explored on the web at Size/Maths). In this sense the standard textbook diagram is a classic illustration of the Escher-like quality of many of the standard economic diagrams: they show what can be drawn, rather than what is truly feasible according to the theory.

From minnow to market

The exposition above simply describes the situation for a single firm. To derive the market supply curve, we have to aggregate the supply curves of a multitude of producers – just as to complete the derivation of the market demand curve, the demand curves of a multitude of consumers had to be added together. The total market supply curve is derived simply by adding up all the individual marginal cost curves of firms in a competitive market. Since each individual firm's marginal cost curve is upward sloping, the market supply curve is also upward sloping: we get the classic market supply curve shown in Figure 7.

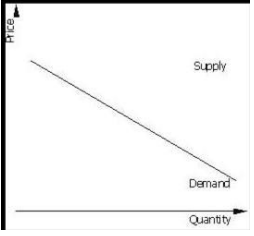


Figure 7: The upward-sloping supply curve is derived by aggregating the marginal cost curves of numerous competitive firms

Things don't add up

There is no doubt that the economic analysis of production has great superficial appeal – sufficient to explain much of the fealty which neoclassical economists swear to their vision of the market. But at a deeper level, the argument is fundamentally flawed – as Piero Sraffa first pointed out in 1926.

The crux of Sraffa's critique was that 'the law of diminishing marginal returns' will not apply in general in an industrial economy. Instead, Sraffa argues that the common position would be constant marginal returns, and therefore horizontal (rather than rising) marginal costs.

Sraffa's argument constitutes a fundamental a critique of economic theory, since diminishing marginal returns determine everything in the economic theory of production. As the previous section has shown, the output function determines marginal product, which in turn determines marginal cost. With diminishing marginal productivity, the marginal cost of production eventually rises to equal marginal revenue. Since firms seek to maximise profit, and since this equality of (rising) marginal cost to marginal revenue gives you maximum profit, this determines the level of output.

If constant returns are the norm, then the output function instead is a straight line through the origin, just like the total revenue line – though with a different slope. If (as a factory owner would hope) the slope of revenue is greater than the slope of the cost curve, then after a firm had met its fixed costs, it would make a profit from every unit sold: the more units it sold, the greater its profit would be.

As economic theory represents it, this implies a situation like Figure 8. At least in terms of the economic model of production, there would be no limit to the amount a competitive firm would wish to produce, so that economic theory could not explain how firms (in a competitive industry) decided how much to produce. In fact, according to economic theory, each firm would want to produce an infinite amount.

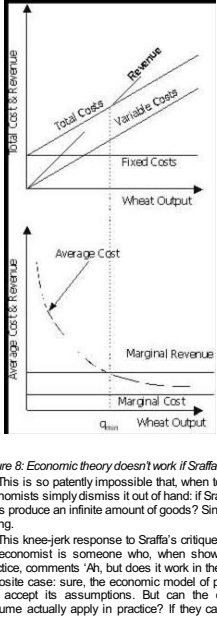


Figure 8: Economic theory doesn't work if Sraffa is right

This is so patently impossible that, when told of Sraffa's critique, many economists simply dismiss it out of hand: if Sraffa was right, then why don't firms produce an infinite amount of goods? Since they don't, Sraffa must be wrong.

This knee-jerk response to Sraffa's critique brings to mind the joke that an economist is someone who, when shown that something works in practice, comments 'Ah, but does it work in theory?'. Sraffa instead put the opposite case: sure, the economic model of production works in theory, if you accept its assumptions. But can the conditions that economists assume actually apply in practice? If they can't, then regardless of how watertight the economic theory of production might be given its assumptions, it will be irrelevant in practice. It therefore should not be used as the theory of production, because above all else, such a theory must be realistic.

Sraffa's argument focussed upon the economic assumptions that there were 'factors of production' which were fixed in the short run, and that supply and demand were independent of each other. He argued that these two assumptions could not be fulfilled simultaneously. In circumstances where it was valid to say that some factor of production was fixed in the short run, supply and demand would not be independent, so that every point on the supply curve would be associated with a different demand curve. On the other hand, in circumstances where supply and demand could justifiably be treated as independent, then in general it would be impossible for any factor of production to be fixed. Hence the marginal costs of production would be constant.

Sraffa began by noting that the preceding classical school of economics also had a 'law of diminishing marginal returns'. However, for the classical school, it was not part of price theory, but part of the theory of income distribution. Its application was largely restricted to the explanation of rent.

The classical argument was that farming would first be done on the best land available, and only when this land was fully utilised would land of a lesser quality be used. Thus, as population grew, progressively poorer land would be brought into use. This poorer land would produce a lower yield per acre than the better land. Diminishing marginal returns therefore applied, but they occurred because the quality of land used fell – not because of any relationship between 'fixed' and 'variable' factors of production.

Sraffa argued that the neoclassical theory of diminishing marginal productivity was based on an inappropriate application of this concept in the context of their model of a competitive economy, where the model assumed that all firms were so small relative to the market that they could not influence the price for their commodity, and that factors of production were homogeneous. In the neoclassical model therefore, falling quality of inputs couldn't explain diminishing marginal productivity. Instead, productivity could only fall because the ratio of 'variable factors of production' to fixed factors exceeded some optimal level.

The question then arises of when is it valid to regard a given factor of production – say, land – as fixed. Sraffa said that this was a valid

assumption when industries were defined very broadly, but this then contradicted the assumption that demand and supply are independent.

Sraffa's broad arrow

If we take the broadest possible definition of an industry – say, agriculture – then it is valid to treat factors it uses heavily (such as land) as fixed. Since additional land can only be obtained by converting land from other uses (such as manufacturing or tourism), it is clearly difficult to increase that factor in the short run. The 'agriculture industry' will therefore suffer from diminishing returns, as predicted.

However, such a broadly defined industry is so big that changes in its output must affect other industries. In particular, an attempt to increase agricultural output will affect the price of the chief variable input – labour – as it takes workers away from other industries (and it will also affect the price of the 'fixed' input).

This might appear to strengthen the case for diminishing returns – since inputs are becoming more expensive as well as less productive. However, it also undermines two other crucial parts of the model: the assumption that demand for and supply of a commodity are independent, and the proposition that one market can be studied in isolation from all other markets.

Instead, if increasing the supply of agriculture changes the relative prices of land and labour, then it will also change the distribution of income. As we saw in Chapter 2, changing the distribution of income changes the demand curve. There will therefore be a different demand curve for every different position along the supply curve for agriculture. This makes it impossible to draw independent demand and supply curves that intersect in just one place. As Sraffa expressed it:

> If in the production of a particular commodity a considerable part of a factor is employed, the total amount of which is fixed or can be increased only at a more than proportional cost, a small increase in the production of the commodity will necessitate a more intense utilisation of that factor, and this will affect in the same manner the cost of the commodity in question and the cost of the other commodities into the production of which that factor enters; and since commodities into the production of which a common special factor enters are frequently, to a certain extent, substitutes for one another the modification in their price will not be without appreciable effects upon demand in the industry concerned. (Sraffa 1926)

These non-negligible impacts upon demand mean that the demand curve for this 'industry' will shift with every movement along its supply curve. It is therefore not legitimate to draw independent demand and supply curves, since factors that alter supply will also alter demand. Supply and demand will therefore intersect in multiple locations, as shown in Figure 9, and it is impossible to say which price or quantity will prevail.

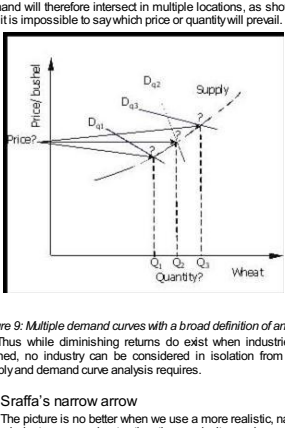


Figure 9: Multiple demand curves with a broad definition of an industry

Thus while diminishing returns do exist when industries are broadly defined, no industry can be considered in isolation from all others, as supply and demand curve analysis requires.

Sraffa's narrow arrow

The picture is no better when we use a more realistic, narrow definition of an industry – say, wheat rather than agriculture – because here Sraffa argues that, in general, diminishing returns are unlikely to exist. This is because the assumption that supply and demand are independent is now reasonable, but the assumption that some factor of production is fixed is not.

While economists assume that production occurs in a period of time

during which it is impossible to vary one factor of production, Sraffa argues that in the real world, firms and industries will normally be able to vary all factors of production fairly easily. This is because these additional inputs can be taken from other industries, or garnered from stocks of under-utilised resources. That is, if there is an increased demand for wheat, then rather than farming a given quantity of land more intensively, farmers will instead convert some land from another crop – say, barley – to wheat. Or they will convert some of their own land which is currently lying fallow to wheat production. Or farmers who currently grow a different crop will convert to wheat. As Sraffa expressed it:

> If we next take an industry which employs only a small part of the 'constant factor' (which appears more appropriate for the study of the particular equilibrium of a single industry), we find that a (small) increase in its production is generally met much more by drawing 'marginal doses' of the constant factor from other industries than by intensifying its own utilisation of it; thus the increase in cost will be practically negligible (Sraffa 1926)

In all cases, the ratio of one factor of production to any other will remain relatively constant, while the total amount of resources devoted to production will rise. This results in a straight-line output function, as shown earlier in Figure 8.

Since the shape of the total, average and marginal cost curves are entirely a product of the shape of the output curve, a straight line output curve results in constant marginal costs, and falling average costs. To put Sraffa's argument in diagrammatic form, a competitive firm in a normal industry would look, not like the picture drawn in Figure 7, but like that in the bottom half of Figure 8.

With this picture of a firm, every additional sale after the point qmin results in more profit: sales are not limited by rising cost, as in economic theory. Instead, costs for a firm are likely to be constant (or even falling) within the normal range of output. If economists are correct that a competitive firm faces a horizontal demand curve (and you might guess that I'm going to dispute this later), then according to economic theory, the firm would be trying to produce an infinite output.

Irrational managers

Sraffa's broad and narrow critiques accept that, if a firm's output was actually constrained by a fixed resource, then its output function would look like that shown in Figure 1. Output would at first rise at an accelerating rate, as the productivity of additional variable inputs rose; then the rate of growth of output would reach a peak as the maximum level of productivity was reached, after which output would still rise, but at a diminishing rate. Finally, when even more variable inputs were added, total output would actually start to fall. In the vernacular of economics, the firm would at first experience increasing marginal productivity, then diminishing marginal productivity, and finally negative marginal productivity.

However, Sraffa disputes even this proposition. He instead argues that a firm is likely to produce at maximum productivity right up until the point at which diminishing marginal productivity sets in. Any other pattern, he argues, shows that the firm is behaving irrationally.

His argument is probably best illustrated with an analogy. Imagine that you have a franchise to supply ice-creams to a football stadium, and that the franchise lets you determine where patrons are seated. If you have a small crowd one night – say one quarter of capacity – would you spread the patrons evenly over the whole stadium, so that each patron was surrounded by several empty seats?

Of course not! This arrangement would simply force your staff to walk further to make a sale. Instead, you'd leave much of the ground empty, thus minimising the work your staff had to do to sell the ice-creams. There's no sense in using every last inch of your 'fixed resource' (the stadium) if demand is less than capacity.

Sraffa argued that the same logic applied to a farm, or to a factory. If a variable input displayed increasing marginal returns at some scale of output, then the sensible thing for the farmer or factory owner to do is leave some of the fixed resource idle, and work the variable input to maximum efficiency on part only of the fixed resource.

To give a numerical example, consider a wheat farm of 100 hectares, where one worker per hectare produces an output of 1 bushel per hectare, 2 workers per hectare produces 3 bushels, 3 per hectare produces 6 bushels, 4 per hectare produces 10 bushels, and 5 workers per hectare produces 12 bushels.

According to economists, if a farmer had 100 workers, he would spread them out 1 per hectare to produce 100 bushels of wheat. But, according to Sraffa, the farmer would instead leave 75 hectares of the farm idle, and work 25 hectares with the 100 workers to produce an output of 250 bushels. The farmer that behaves as Sraffa predicts comes out 150 bushels ahead of any farmer who behaves as economists expect.

Similarly, economic theory implies that a farmer with 200 workers would spread them over the farm's 100 hectares, to produce an output of 300

bushels. But Sraffa says that a sensible farmer would instead leave 50 hectares fallow, work the other 50 at 4 workers per hectare, and produce an output of 500 bushels. One again, a 'Sraffian' farmer is ahead of an 'economic' one, this time by 200 bushels.

The same pattern continues right up until the point at which 400 workers are employed, when finally diminishing marginal productivity sets in. A farm will produce more output by using less than all of the fixed input, up until this point.

This situation is depicted in Figure 10, where the shaded area indicates the output forgone by a firm that uses all of its fixed resource all of the time, rather than leaving some of it idle until it reaches maximum productivity. Clearly, a 'rational' firm with a fixed resource would leave some of it idle when operating below its optimum scale. Firms will therefore have straight line marginal cost curves below the level of maximum productivity.

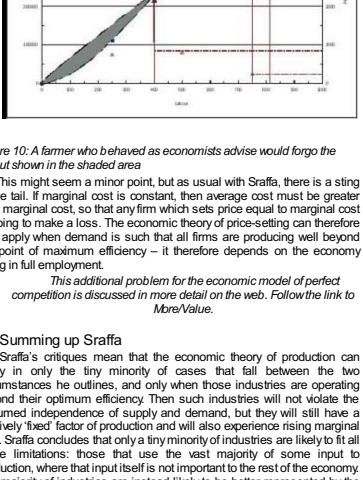


Figure 10: A farmer who behaved as economists advise would forgo the output shown in the shaded area

This might seem a minor point, but as usual with Sraffa, there is a sting in the tail. If marginal cost is constant, then average cost must be greater than marginal cost, so that any firm which sets price equal to marginal cost is going to make a loss. The economic theory of price-setting can therefore only apply when demand is such that all firms are producing well beyond the point of maximum efficiency – it therefore depends on the economy being in full employment.

This additional problem for the economic model of perfect competition is discussed in more detail on the web. Follow the link to More/Value.

Summing up Sraffa

Sraffa's critiques mean that the economic theory of production can apply in only the tiny minority of cases that fall between the two circumstances he outlines, and only when those industries are operating beyond their optimum efficiency. Then such industries will not violate the assumed independence of supply and demand, but they will still have a relatively 'fixed' factor of production and will also experience rising marginal cost. Sraffa concludes that only a tiny minority of industries are likely to fit all these limitations: those that use the vast majority of some input to production, where that input itself is not important to the rest of the economy. The majority of industries are instead likely to be better represented by the classical theory, which saw prices as being determined exclusively by costs, while demand set the quantity sold. This results in an industry-level picture like that shown in Figure 11. As Sraffa put it:

> Reduced within such restricted limits, the supply schedule with variable costs cannot claim to be a general conception applicable to normal industries; it can prove a useful instrument only in regard to such exceptional industries as can reasonably satisfy its conditions. In normal cases the cost of production of commodities produced competitivelymust be regarded as constant in respect of small variation in the quantity produced. And so, as a simple way of approaching the problem of competitive value, the old and now obsolete theory which makes it dependent on the cost of production alone appears to hold its ground as the best available. (Sraffa 1926)

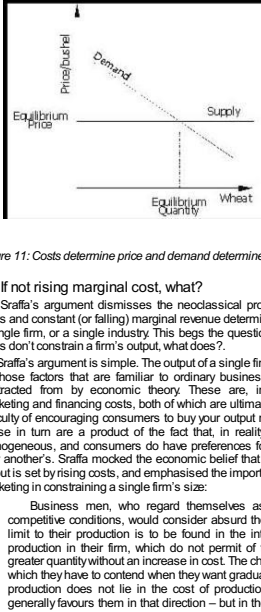


Figure 11: Costs determine price and demand determines quantity

If not rising marginal cost, what?

Sraffa's argument dismisses the neoclassical proposition that rising costs and constant (or falling) marginal revenue determines the output from a single firm, or a single industry. This begs the question that if increasing costs don't constrain a firm's output, what does?.

Sraffa's argument is simple. The output of a single firm is constrained by all those factors that are familiar to ordinary businessmen, but that are abstracted from by economic theory. These are, in particular, rising marketing and financing costs, both of which are ultimately a product of the difficulty of encouraging consumers to buy your output rather than a rival's. These in turn are a product of the fact that, in reality, products are not homogeneous, and consumers do have preferences for one firm's output over another's. Sraffa mocked the economic belief that the limit to a firm's output is set by rising costs, and emphasised the importance of finance and marketing in constraining a single firm's size:

> Business men, who regard themselves as being subject to competitive conditions, would consider absurd the assertion that the limit to their production is to be found in the internal conditions of production in their firm, which do not permit of the production of a greater quantity without an increase in cost. The chief obstacle against which they have to contend when they want gradually to increase their production does not lie in the cost of production – which, indeed, generally favours them in that direction – but in the difficulty of selling the larger quantity of goods without reducing the price, or without having to face increased marketing expenses. (Sraffa 1926)

Economics assumes this real world answer away by assuming that products are homogeneous, consumers are indifferent between the outputs of different firms and decide their purchases solely on the basis of price, that there are no transportation costs, etc. In such a world, no-one needs marketing, because consumers already know everything, and only price (which consumers already know) distinguishes one firm's output from another. But Sraffa says that these postulates are the exception to the rule which applies in reality.

In most industries, products are heterogeneous, consumers do not know everything, and they consider other aspects of a product apart from price. Even where products are homogeneous, transportation costs can act to give a single firm an effective local monopoly. As a result, even the concept of a competitive market – in which all firms are price-takers – is itself suspect. Instead, most firms will to varying degrees act like monopolists – who, according to economic theory, face a downward-sloping demand curve.

Each firm has a product that may fit within some broad category – such as, for example, passenger cars – but which is qualitatively distinguished from its rivals in a fashion that matters to a particular subset of buyers. The firm attempts to manipulate the demand for its product, but faces prohibitive costs in any attempt to completely eliminate their competitors and thus take over the entire industry. Not only must the firm persuade a different niche market to buy its product – to convince Porsche buyers to buy Volvos, for example – it must also convince investors and banks that the expense of building a factory big enough to produce for both market niches is worth the risk. Therefore, with the difficulty of marketing beyond your product's niche goes the problem of raising finance:

> The limited credit of many firms, which does not permit any one of them to obtain more than a limited amount of capital at the current rate of interest, is often a direct consequence of its being known that a given firm is unable to increase its sales outside its own particular market without incurring heavy marketing expenses. (Sraffa 1926)

Economic theory also can't be saved by simply adding marketing costs to the cost of production, and thus generating a rising marginal cost curve. As Sraffa pointed out, there are at least three flaws with this. Firstly, it is a

distortion of the truth – marketing is not a cost of production, but a cost of distribution. Secondly, it is inconsistent with the underlying economic premise that marginal cost rises because of diminishing marginal productivity. Thirdly, it is implausible in the context of the economic theory of the firm. There is no point in 'saving' the concept of a rising marginal cost curve by introducing marketing costs, since this requires acknowledging that one firm's product differs from another. If products differ from one firm to another, then products are no longer homogeneous, which is an essential assumption of the theory of perfect competition. It is far more legitimate to treat marketing as a cost of distribution, whose object is to alter the demand faced by an individual firm.

If Sraffa is correct, then that most popular and pervasive image of economic theory, a falling demand curve and rising supply curve intersecting to jointly determine the equilibrium price, is an illusion. Rather than firms producing at the point where marginal cost equals marginal revenue, the marginal revenue of the final unit sold will normally be substantially greater than the marginal cost of producing it, and output will be constrained, not by marginal cost, but by the cost and difficulty of expanding sales at the expense of sales by competitors.[26] At best, the situation for each market will be similar to that shown in Figure 10 – if we are generous and ignore the result shown in Chapter 2 that the demand curve can't be drawn as smoothly downward sloping.

Sraffa's alternative

However, Sraffa was not satisfied with this revised picture, which is still dominated by intersecting marginal revenue and marginal cost curves. He instead expressed a preference for a more realistic model, which focussed upon those issues that are most relevant to actual businesses.

To put Sraffa's preferred analysis visually, the situation is similar to that shown in Figure 12. The firm faces falling average costs as its large fixed costs are amortised over a larger volume of sales, and as its variable costs either remain constant or fall with higher output.[27] It will have a target level of output which it tries to exceed, and a target markup which it tries to maintain. The size of the firm is constrained by the size of its niche within the given market, and the difficulty of raising finance for a much larger scale of operation.

The margin between costs of production and target sale price will be set by the degree of product differentiation within the industry, competitive pressures and general market conditions. Each firm will endeavour to sell as much output as it can, but the level of output will be constrained by the size of the firm's market niche and the marketing efforts of rivals.

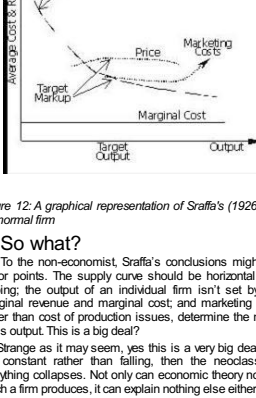


Figure 12: A graphical representation of Sraffa's (1926) preferred model of the normal firm

So what?

To the non-economist, Sraffa's conclusions might still look like fairly minor points. The supply curve should be horizontal rather than upward sloping; the output of an individual firm isn't set by the intersection of marginal revenue and marginal cost; and marketing and finance issues, rather than cost of production issues, determine the maximum scale of a firm's output. This is a big deal?

Strange as it may seem, yes this is a very big deal. If marginal returns are constant rather than falling, then the neoclassical explanation of everything collapses. Not only can economic theory no longer explain how much a firm produces, it can explain nothing else either.

Take, for example, the economic theory of employment and wage determination (discussed in more detail in Chapter 5). The theory asserts that the real wage is equivalent to the marginal product of labour. The argument goes that each employer takes the wage level as given, since with

competitive markets no employer can affect the price of his inputs. An employer will employ an additional worker if the amount the worker adds to output – the worker's marginal product – exceeds the real wage. The employer stops employing workers once the marginal product of the last one employed has fallen to the same level as the real wage.

The picture is as shown in Figure 13, and it explains the economic predilection for blaming everything on wages being too high – neoclassical economics can be summed up, as Galbraith once remarked, in the twin propositions that the poor don't work hard enough because they're paid too much, and the rich don't work hard enough because they're not paid enough (Galbraith 1997). The output of the firm, in the top diagram of Figure 13, is subject to diminishing marginal returns, and thus marginal product – the slope of the output curve, shown in the bottom diagram – falls as output increases. The real wage, in the bottom diagram, is unaffected by the output level of the firm. The firm will keep on hiring workers until the marginal product of the last worker equals the real wage. This occurs at point C.

Since the rational employer stops hiring at point C, the real wage – which the employer takes as given – determines how many workers this firm employs. Since employment in turn determines output, the real wage determines the level of output. If society desires a higher level of employment and output, then the only way to get it is to reduce the real wage (and the logical limit of this argument is that output will reach its maximum when the real wage equals zero). The real wage in turn is determined by the willingness of workers to work – to forgo leisure for income – so that ultimately the level of employment is determined by workers alone.

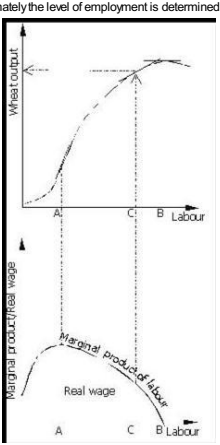


Figure 13: The economic theory of income distribution argues that the wage equals the marginal product of labour

If in fact the output to employment relationship is relatively constant, then the neoclassical explanation for employment and output determination collapses. With a flat production function, the marginal product of labour will be constant, and it will never intersect the real wage. The output of the firm then can't be explained by the cost of employing labour. The picture is as shown in Figure 14, and neoclassical economics simply cannot explain anything: neither the level of employment, nor output, nor, ultimately, what determines the real wage.

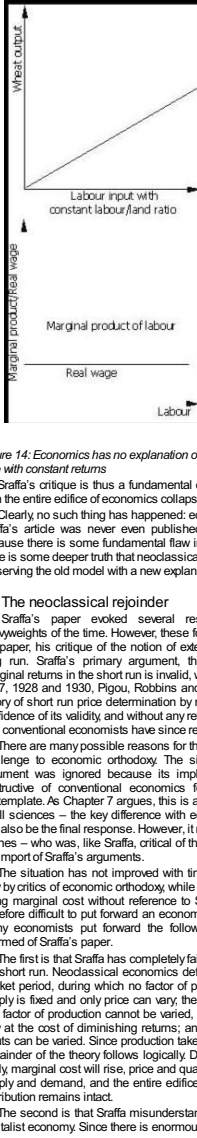


Figure 14: Economics has no explanation of wage determination or anything else with constant returns

Sraffa's critique is thus a fundamental one: if his argument is accepted then the entire edifice of economics collapses.

Clearly, no such thing has happened: economics has continued on as if Sraffa's article was never even published. One might hope that this is because there is some fundamental flaw in Sraffa's argument, or because there is some deeper truth that neoclassical economics discovered to justify preserving the old model with a new explanation. Sadly, neither is the case.

The neoclassical rejoinder

Sraffa's paper evoked several responses from the economic heavyweights of the time. However, these focussed upon another aspect of the paper, his critique of the notion of external economies of scale in the long run. Sraffa's primary argument, that the concept of diminishing marginal returns in the short run is invalid, was ignored – so much so that in 1927, 1928 and 1930, Pigou, Robbins and Harrod respectively set out the theory of short run price determination by rising marginal cost in complete confidence of its validity, and without any reference to Sraffa's paper. Few, if any, conventional economists have since referred to Sraffa's paper.

There are many possible reasons for this complete neglect of a serious challenge to economic orthodoxy. The simplest explanation is that the argument was ignored because its implications, if accepted, were too destructive of conventional economics for conventional economists to contemplate. As Chapter 7 argues, this is a not uncommon initial response in all sciences – the key difference with economic science being that this can also be the final response. However, it must be acknowledged that even Keynes – who was, like Sraffa, critical of the mainstream – failed to realise the import of Sraffa's arguments.

The situation has not improved with time. Sraffa's paper is today cited only by critics of economic orthodoxy, while the textbooks teach the theory of rising marginal cost without reference to Sraffa's counter-arguments. It is therefore difficult to put forward an economic response to Sraffa. However, many economists put forward the following arguments when they are informed of Sraffa's paper.

The first is that Sraffa has completely failed to understand the concept of the short run. Neoclassical economics defines three concepts of time: the market period, during which no factor of production can be varied, so that supply is fixed and only price can vary; the short run, during which at least one factor of production cannot be varied, so that output can be varied but only at the cost of diminishing returns; and the long run, during which all inputs can be varied. Since production takes place during the short run, the remainder of the theory follows logically. Diminishing marginal returns will apply, marginal cost will rise, price and quantity will be jointly determined by supply and demand, and the entire edifice of the theory of production and distribution remains intact.

The second is that Sraffa misunderstands the nature of production in a capitalist economy. Since there is enormous pressure to be competitive, no

firm can survive long with excess capacity. Therefore competition will drive all firms towards full capacity, and in this realm diminishing returns will apply.

Time and the short run

As Chapter 8 points out, time – or rather, the absence of time in its analysis – is one of the fundamental weaknesses of conventional economics. It is therefore somewhat ironic that economists defend their theory from attack by appealing to the importance of time. However, far from helping to defend economic theory from criticism, the proper analysis of time highlights a critical weakness.

A firm's revenue and costs clearly vary over time, as well as varying as the firm changes its level of output at any one point in time. The economic rule that (in the context of diminishing returns) 'profit is maximised where marginal cost equals marginal revenue' is derived by 'holding time constant' and thus describing revenue and cost as simply a function of the quantity produced. The gap between revenue and cost is widest where marginal cost equals marginal revenue.

But in fact this rule applies only 'when time stands still' – which time never does. Not even an economist can make time stand still (though some victims of economics lectures might dispute that!). Similarly, the rule tells you how to maximise profit with respect to quantity, but real businessmen are more interested in maximising profit over both time and output.

It is possible to consider profit as a function of both time and quantity, as opposed to the economic approach of dividing time into artificial segments, by explicitly acknowledging that profit is a function of both time and quantity (which the firm can vary at any point in time, and that will also change – and hopefully grow – over time).[28] Profit therefore depends both on the amount a firm produces, and the historical time during which it produces.

Using a rule of mathematics, we can then decompose the change in profit into the contribution due to the progress of time, and the contribution due to changes in quantity (which will also change over time). This results in the formula:

Change in profit equals
change in profit due to change in time
multiplied by the change in time, plus
change in profit due to change in quantity
multiplied by the change in quantity

This formula tells us how big a change in profit will be, so if a firm wants to maximise its profit, it wants this number to be as big as possible.

Change in profit due to change in quantity is the same thing as 'marginal revenue minus marginal cost'. Economic theory argues that profit is maximised when marginal revenue equals marginal cost, so that if you follow the economic profit maximisation rule, you would deliberately set this quantity to zero. Since you get zero when you multiply any number by zero, following this economic rule sets the second half of the formula (change in profit due to change in quantity multiplied by the change in quantity) to zero.

Therefore, economic theory tells us that the change in profit will be maximised when we eliminate the contribution that changes in quantity make to changes in profit. Change in profit is thus reduced simply to the first half of the formula, where changes due to time alone determine the change in profit. But economic theory has given us no advice about how to make change in profit due to change in time as big as possible.

What's going on? Suddenly, advice that previously seemed sensible (before we considered Sraffa's critique of the notion of a fixed factor of production) looks obviously absurd. Clearly something is wrong: but what? Let's go back to that formula, which is true by definition, and see what it tells us to do.

If the firm's output is growing over time, then the term change in quantity will be positive. Setting marginal revenue equal to marginal cost means multiplying this positive number by zero – which results in a smaller increase in profit than if marginal revenue exceeds marginal cost. With rising sales, you will get a higher profit if 'change in profit due to change in quantity' is also positive, which requires that marginal revenue be greater than marginal cost. Thus a careful consideration of time argues that a firm should ensure that its marginal revenue is greater than its marginal cost. This is the position which Sraffa argued actually applies in reality, so that the proper consideration of time strengthens Sraffa's critique, rather than weakening it.

The economic rule about how to maximise profit is therefore correct only if the quantity produced never changes.[29] This argument might appear paradoxical, especially if you have been conditioned to thinking like an economist.[30]

An analogy might help you interpret it. Imagine you have a formula which describes how much fuel your car uses at any given speed, and you want to work out the most economical speed at which to drive. What you need to do is to work out the lowest rate at which to consume petrol per unit of distance

travelled per second. If instead you first work out the most economical speed at which to travel, the answer to this first question will be zero miles per hour – because at this speed you consume the lowest possible amount of petrol per unit of time, zero. This is an accurate but useless answer, since you're not interested in staying put. If you want to work out the speed that minimises petrol consumed but still gets you to your destination, you have to handle both problems simultaneously.

The economic theory of the firm ignores time, in the same way that the wrong answer to the 'most economical speed at which to travel' question ignores distance. But time is an essential aspect of economic behaviour, in the same sense that distance is an essential aspect of travel. The economic policy for profit maximisation is derived by following the economic equivalent of finding the cheapest speed at which to travel first, and then multiplying by the distance travelled. It is thus a kindred spirit to the advice that the cheapest way to get from point A to point B is to travel at zero miles per hour.

There is also an economic way of interpreting this apparent paradox that advice which appears sound when you ignore (or compartmentalise) time becomes absurd when you take time into account. This is that, by ignoring time in its analysis of the firm, economic theory ignores some of the most important issues facing a firm. Its 'static' emphasis upon maximising profit 'now' ignores the fact that, to survive, a firm must also grow over time. To grow it must invest and develop new products, and this takes energy and resources. If instead it devotes all its resources to maximising profit now, then it will not have any energy or resources left to devote to investment and new product development.

If we try to interpret economic theory in the context of historical time, then what the theory is attempting to do is work out the ideal level of output of a product for all time. But in the real world there is no such level of output. The appropriate number of motor cars to produce in 1900 was quite different from the appropriate number to produce in 2000.

This is how something that once looked so right (before Sraffa's critique of the concept of a fixed factor of production) looks so absurd now. The formula discussed above explicitly takes time into account, and is therefore dynamic, while the economic theory of the firm is static: it ignores time, and is therefore only relevant in a world in which time does not matter. But time clearly does matter in our world, and what is right in a static setting is wrong in a dynamic one.

This is one of the many instances of the phenomenon I mentioned in Chapter 1, that advice derived from static reasoning, which ignores time, is often categorically opposed to advice derived from dynamic analysis, which takes time into account. Since the economy is fundamentally dynamic, static analysis is therefore normally dangerously wrong. I explore these issues in more depth in Chapter 8.

The flaws in economic reasoning pointed out in this chapter and Chapter 4 have a very direct impact on public policy in the area of the pricing of public services. Because economists believe that competitive industries set price equal to marginal cost, economists normally pressure public utilities to price their services at 'marginal cost'. Since the marginal costs of production are normally constant and well below the average costs, this policy will normally result in public utilities making a loss. This is likely to mean that public utilities are not able to finance the investment they need in order to maintain the quality of services over time. This dilemma in turn interacts with the pressure that economists also apply to privatise public assets, and to let individuals 'opt out' of the public provision of essential services. The end result, as Galbraith so eloquently put it, is "private affluence and public squalor".

Ironically, economic theory also makes economists essentially 'anti-capitalist', in that they deride real businesses for pricing by a markup on cost, when theory tells them that prices should be set at the much lower level of marginal cost. Industrialists who have to cope with these attitudes in their dealings with government-employed economists are often amongst the greatest closet anti-economists of all. Maybe it's time for them to come out of the closet.

Competition and capacity

The argument that competition would drive all firms to use their fixed capital at full capacity does look convincing at first glance, but deeper thought reaches precisely the opposite conclusion. A firm with no spare capacity has no flexibility to take advantage of sudden, unexpected changes in the market, and it also has to consider building a new factory as soon as its output grows. The economist Janos Kornai has explored this issue in detail, and concluded that excess capacity is essential for survival in a market economy.

This issue is examined in detail on the web. Follow the link to More/Value.

But wait, there's more

Sraffa's critique was entirely based upon an appeal to logic; a defence which might appear to be open to economic theory was, of course, that the facts supported them rather than Sraffa. However, empirical research has

shown that the vast majority of firms operate with constant or falling marginal costs, and set their prices without reference to the economic concepts of marginal cost or marginal revenue. For example, when factory managers were asked to choose between graphs which illustrated different ways in which costs might change with output, almost 95% nominated figures which did not illustrate rising marginal costs (see Table 2).

	By Firms	By Products
Supports MC=MR	18	62
Contradicts MC=MR	316	1020
Per Cent supporting MC=MR	5.4	5.7

Table 1: Empirical research on nature of cost curves

Numerous other empirical studies have found that the data supports Sraffa: most firms produce under conditions of constant or falling marginal costs, and the conventional economic arguments about production are simply wrong in the real world.

This issue is examined in detail on the web. Follow the link to More/Value.

A totem in tatters

While economic theory is thus fallacious, it is still feasible that, in some instances, price will rise as output rises. Feasible 'real-world' reasons for this include the inflexibility of supply in some markets across some timeframes (something economics attempts to deal with by its concept of the market period, as opposed to the short run theory debunked in this chapter), firms exploiting high demand to set higher margins, and, in some circumstances, wage demands rising during periods of high employment.[31] But the economic attempt to link higher prices directly to declining productivity is a failure.

This of itself would not be catastrophic, were it not for the extent to which diminishing marginal productivity permeates economic theory. It is a foundation stone which, when it is withdrawn, brings down virtually everything else with it. Sraffa's critique thus provides one more illustration of the remarkable fragility of this outwardly confident social theory we call economics. Economics is not the Emperor of the social sciences, but the Humpty Dumpty.

Just like Humpty Dumpty after his fall, it is impossible to reconstruct the totemic supply and demand diagram of economic theory after the criticisms outlined in this and the preceding chapter. According to the Sonnenshien-Mantel-Debreu conditions, a market demand curve aggregated in accordance with economic theory can have any shape at all – except, arguably, the smooth, downward-sloping curve which economists always draw in their textbooks. According to Sraffa's critique (and empirical research into costs and pricing), the supply curve is at best flat, and probably downward sloping, because the economic construct of a fixed factor of production does not apply in the real world. All that economists can legitimately construct is the mangled creature shown in Figure 15. In place of the strong and definitive totem which economists hold aloft for their students, we have a coathanger wrapped around a stick.[32]

No economist would proudly salute this flag, and this symbolic failure is not the end of the problems for economics. Virtually every other concept in economic theory depends on the notions of diminishing marginal productivity for firms on the one hand, and diminishing marginal utility for the community on the other. If both these foundations are unsound, then almost nothing else in economic theory remains standing. Without diminishing marginal productivity, economics cannot explain how a firm decides how much to produce. This alone invalidates their analysis of market structures and income distribution. Without a community utility map, everything from the analysis of optimum output levels to the theory of international trade collapses.

This is not the end of the story either. Both chapters 2 and 3 implicitly accepted that a supply curve and a demand curve are all that is needed to work out price and quantity. But the next chapter shows that this is simply not possible: the economic argument that price is set by the intersection of supply and demand cannot possibly be true.

4. Size does matter

Why the economic argument against monopolies is invalid

The historical record gives many reasons to be critical of monopolies. Some of today's most respectable names accumulated fortunes in the 19th century by grossly unethical means, and held on to their economic wealth by the abuse of the many forms of power that wealth can confer. This record of 19th century corporate lawlessness is the major explanation for America's distinctive anti-trust laws, which enable the judiciary in this bastion of capitalism to dismantle corporations that have acquired the position of ultimate market power.

Economists have reduced this plethora of reasons to be critical of monopolies to just one: the size of a monopolist relative to the size of the market. According to economists, monopolies are worse than the alternative of competitive markets simply because the absence of competitors allows a monopoly to set price above marginal cost. Perfectly competitive industries, on the other hand, set price equal to marginal cost.[33]

Economists argue that monopolies are therefore necessarily bad, and that small, competitive firms are necessarily good, because monopolies produce a lower output for a higher price, thus restricting the supply of their products and exploiting consumers by over pricing. However, the 'proof' of this argument makes the fundamental mathematical mistake of equating a very small quantity with zero. When this mistake is corrected, the economic argument against monopolies collapses, as does the economic theory of competition.

The kernel

Try this party trick: convince someone that the world is flat, starting from the premise that it is a sphere.

The argument is simple. If you take a small enough segment of the world – say, the two feet your victim is standing on – then the curvature of that segment is so small that it is, for all intents and purposes, flat. Then consider the segment you're standing on – it is also so small that it is effectively flat.

Next, consider the angle between the two segments: it too will be so small that it is effectively zero. So these two small segments are effectively flat.

Finally, extrapolate your argument from these two tiny segments and the angle between them up to the level of the entire globe. If you consider the segment your victim occupies and the segment behind him/her, that pair is also effectively flat. Keep on going, and the entire world is flat.

The fallacy in the argument, clearly, is that while it will do as an approximation to treat your immediate surroundings as effectively flat, it will not do to ignore those small but imperceptible angles if you move from the scale of one or two segments to the entire globe.

Yet this fallacy lies at the heart of the economic preference for small, competitive firms over large monopolistic ones. At crucial stages of the economic argument, an imperceptibly small quantity is treated as zero, and then all these zeros are added up to yield zero at the scale of an entire market. This is intellectually and mathematically unsound. When the correct position is imposed – that something which is extremely small is nonetheless not zero – the economic argument against monopolies and in favour of small, competitive firms collapses.

Oh, and if your party trick convinces your victim? Then the person you're talking to is either stoned, or an economist.

The roadmap

In this chapter I outline the neoclassical analysis of monopolies on the one hand, and 'perfect competition' on the other, and point out that the sole difference between them is that a monopolist is shown to face falling marginal revenue, whereas the competitive firm faces constant marginal revenue. I then show that this proposition leads to logical fallacies when you disallow the economic sleight of hand of treating a small quantity as identically equal to zero.

The result is a minefield of logical conundrums: a quantity that economists assume is zero actually has to be minus one; firms that are allegedly profit maximisers must produce more than the amount which maximises profits; zero amounts at the individual level must somehow aggregate to positive amounts at the aggregate; and worst of all, if didactic comparisons are to be made between perfect competition and monopoly, then the key economic concept of diminishing marginal productivity (see Chapter 3) cannot apply. Finally, it is shown (on the web) that to definitively compare competitive firms to a monopoly, marginal costs must be constant; and that the position which economists argue maximises profits for the individual competitive firm does not in fact maximise its profits (the formal

proofs of these results are on the web at Maths/Size).

Economic perfection

Pejorative expressions abound in economics, despite its claim to be a value-free science, and 'perfect competition' is possibly the most value-laden of all. To economists, however, the word 'perfect' has a very precise meaning: it is a market in which the competitively set price equals the marginal cost of production.

This is 'perfect' because, according to economic theory, it achieves the maximum possible gap between community welfare and the cost of providing it. Community welfare is maximised when the marginal benefit from a commodity equals its marginal cost, because then the gap between total benefits to society and the total cost of providing those benefits is as big as it can be.

Given the theory of utility maximisation, the price paid for a commodity represents its marginal contribution to community welfare. Therefore, community welfare is maximised when the marginal cost of production equals the price for which a commodity is sold, and only perfect competition guarantees this outcome. In any other market structure, price exceeds marginal cost.

Unfortunately, this concept of economic perfection relies upon downward-sloping demand curves, diminishing marginal productivity, and a static vision of profit maximisation. As chapters 2 and 3 have shown, these concepts are generally invalid. However, even if we accept, for the sake of argument, that the market demand curve is smoothly downward sloping and represents community welfare, that production is subject to diminishing returns so that the supply curve represents the marginal cost of production, and that static analysis is valid, the economic argument for the superiority of the perfectly competitive market over the monopoly firm is still internally flawed. To establish this, we'll first consider the market form least favoured by economics: monopoly.[34]

Monopoly

A monopoly has the entire market demand curve to itself. Since economists assume that the market demand curve is smoothly downward sloping, the price at which its output can be sold decreases as the quantity it tries to sell grows. A hypothetical example is shown in Table 1. The firm can sell 1 unit of output for $100. Its total revenue is therefore $100, and its 'marginal revenue' – the change in total revenue from zero for zero units sold, to $100 for one unit sold – is also $100.

The price consumers are willing to pay drops smoothly as the quantity supplied rises, so that for an output of 9 units, the sale price has to drop to $99.91. Total revenue for selling nine units would be $899.19. If 10 units were to be sold, the monopolist would have to drop the price per unit by a cent, to $99.90 each. Total revenue would be $999, and marginal revenue – the change in total revenue from $899.19 at the previous output level of 9 units to $999 at the present level of 10 units – would be $99.80.

The same process continues indefinitely, so that if output were 1000 units, then sale price would have to drop to $90. Total revenue would be $90,000, and marginal revenue – the amount of additional revenue added by selling the 1000th unit – would be $80.

Eventually, the point is reached at which any further increase in output requires a price cut which reduces, rather than increases, total revenue. In Table 1, this occurs at an output of 5000 units, where the sale price is $50. The sale of the 5000th unit adds nothing to total revenue, and any increase in sales past this point actually reduces total revenue – marginal revenue has become negative.

Quantity	Price	Total Revenue	Marginal Revenue
1	100	100	100
9	99.91	899.19	99.82
10	99.9	999	99.81
100	99	9900	98
1000	90	90000	80
2000	80	160000	60
3000	70	210000	40
4000	60	240000	20
5000	50	250000	0
6000	40	240000	-20

Table 1: Demand schedule for a hypothetical monopoly

This same argument is shown graphically in Figure 1. Initially a reduction in price will be more than compensated for by an increase in sales: total revenue rises and marginal revenue (the change in revenue for a change in price) is positive. But at some point any further reduction in price results in a less than compensating increase in sales, and total revenue

reaches a maximum (point A in Figure 1).

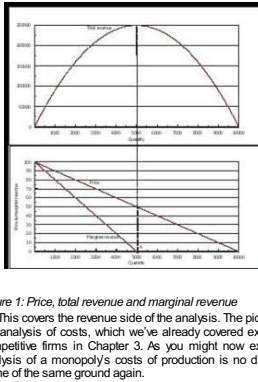


Figure 1: Price, total revenue and marginal revenue

This covers the revenue side of the analysis. The picture is completed by the analysis of costs, which we've already covered exhaustively for small, competitive firms in Chapter 3. As you might now expect, the economic analysis of a monopoly's costs of production is no different; but I'll cover some of the same ground again.

The firm has two types of costs: fixed costs, which apply no matter what the level of output is, and variable costs, which depend directly on how many units are produced. Since diminishing marginal returns are assumed, the monopoly faces rising variable costs of production.

These two costs are shown in the second and third columns of Table 2. Fixed costs are constant at $10,000; average variable costs start at $15 per unit, and initially they fall (the firm experiences rising marginal productivity). As a result, marginal cost falls. However, after a while marginal productivity starts to fall, so that, as shown in Chapter 3, marginal cost begins to rise, as shown in column 4 of Table 2.

Quantity	Fixed Cost	Variable Cost	Marginal Cost
1	10000	15	15
9	10000	135	15
10	10000	150	15
100	10000	1475	15
1000	10000	12861	11
1999	10000	23766	12
2000	10000	23778	12
3000	10000	38250	19
4000	10000	64444	35
5000	10000	113194	64
6000	10000	198000	108

Table 2: Costs for a hypothetical monopoly

Table 3 combines the revenue information from Table 1 with the cost information from Table 2. So long as the last column – showing marginal cost – is smaller than the second last column – showing marginal revenue – then the firm's profit rises. But once marginal revenue and marginal cost are equal, at an output of between 3614 and 3615 units, profit is maximised at between $168,477.51 and $168,477.52. The price the firm charges at this output level equals the market demand price for that output, of between $63.85 and $63.86. Thus to maximise its profits, the firm produces where marginal cost equals marginal revenue, and sells the output at a much higher price.

Quantity	Price	Revenue	Cost	Profit	Av. Cost
1	100.00	100.00	10,015.00	-9,915.00	10,015.0
9	99.91	899.19	10,134.80	-9,235.61	1,126.09
10	99.90	999.00	10,149.75	-9,150.75	1,014.98
99	99.01	9,801.99	11,460.75	-1,658.76	115.77
100	99.00	9,900.00	11,475.26	-1,575.26	114.75
999	90.01	89,919.99	22,849.92	67,070.07	22.87
1000	90.00	90,000.00	22,861.11	67,138.89	22.86
2000	80.00	160,000.00	33,777.78	126,222.22	16.89
3000	70.00	210,000.00	48,250.00	161,750.00	16.08
3600	64.00	230,400.00	61,926.40	168,473.60	17.20
3610	63.90	230,679.00	62,201.85	168,477.15	17.23

3611	63.89	230,706.79	62,229.49	168,477.30	17.23
3612	63.88	230,734.56	62,257.15	168,477.41	17.24
3613	63.87	230,762.31	62,284.83	168,477.48	17.24
3614	63.86	230,790.04	62,312.52	168,477.52	17.24
3615	63.85	230,817.75	62,340.24	168,477.51	17.24
3616	63.84	230,845.44	62,367.97	168,477.47	17.25
3617	63.83	230,873.11	62,395.72	168,477.39	17.25
3618	63.82	230,900.76	62,423.48	168,477.28	17.25

Table 3: Sales and costs determine the level of output which maximises profit

Figure 2 shows the information in Table 3 graphically. The top half of the figure shows total revenue, total cost and total profit (columns 3–5 of Table 3); the bottom half shows the average cost, marginal revenue and marginal cost of producing each unit, as the total volume of production increases (columns 6.8 of Table 3). Average variable costs are 'u-shaped', since they fall while the firm experiences rising marginal productivity, flatten out as diminishing marginal productivity kicks in, and finally rise when marginal productivity has diminished so much that each additional unit costs more to produce than the average to date. Initially, marginal revenue exceeds marginal cost, but as output rises the former falls and the latter rises until, at point Qe in Figure 2, they are equal. After that point, each additional unit produced adds more to cost than it does to revenue, and thus reduces profit. The firm therefore achieves maximum profit by producing an output at which marginal revenue equals marginal cost.

Average price per unit sold at this level of output is Pe in Figure 2, while average cost is Ce (with the 'e' signifying equilibrium). This price exceeds the marginal cost of producing the last unit sold. This means, in economic welfare terms, that the marginal benefit of the last unit sold exceeds the marginal cost of producing it. Society would therefore benefit from an increased level of production, since additional units of output would increase social welfare. But the monopolist has no incentive to produce more: in fact producing any more would reduce her profits. Therefore, according to economists, monopolies reduce social welfare.

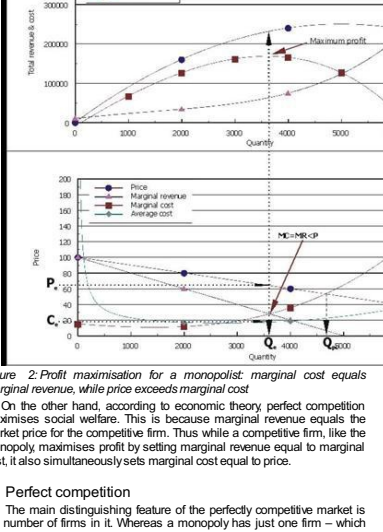


Figure 2: Profit maximisation for a monopolist: marginal cost equals marginal revenue, while price exceeds marginal cost

On the other hand, according to economic theory, perfect competition maximises social welfare. This is because marginal revenue equals the market price for the competitive firm. Thus while a competitive firm, like the monopoly, maximises profit by setting marginal revenue equal to marginal cost, it also simultaneously sets marginal cost equal to price.

Perfect competition

The main distinguishing feature of the perfectly competitive market is the number of firms in it. Whereas a monopoly has just one firm – which therefore has the entire market demand curve to itself – a perfectly competitive market has myriad firms, each competing for a tiny slice of total demand.

In addition, it is assumed that entry into and exit from a competitive industry is 'free', or more accurately not subject to any barriers. Therefore firms outside the industry can move in at any time to take advantage of any above-normal profits if they exist.[35]

All are assumed to produce a product that is homogenous from the consumers' point of view, so that there is no brand loyalty. All firms are therefore 'price-takers': they cannot influence the market price, but instead must take price as given.

The market demand situation is the same as shown for a monopoly in Table 1 and Figure 1 – market demand is still a negative function of price. Therefore, total market revenue will initially be a rising and then a falling function of price, and marginal revenue at the market level will be less than price (because to increase overall sales, the average price must fall).

However, economists argue that for each price-taking firm, marginal revenue and price are identical. The argument is that since they are each so small, no single firm can influence the market price. As a result, if any firm increases its price above the market equilibrium price, it will lose all its customers; while if any firm decreases its price below the market equilibrium, it will suddenly be swamped by all customers for that commodity. Therefore, the firm effectively sees a horizontal demand curve (if this proposition disturbs you, good – because this is the point at which the economic argument starts to take on that 'flat earth' feeling).

Figure 3 shows the situation which economists argue applies to a perfectly competitive firmf Note that the total revenue 'curve' for a perfectly competitive firm is a straight line: increasing its sales has no effect on its price, hence marginal revenue is constant. This gives the model of the perfectly competitive market one key attribute that is missing from the model of a monopoly: a 'supply curve' can be derived.

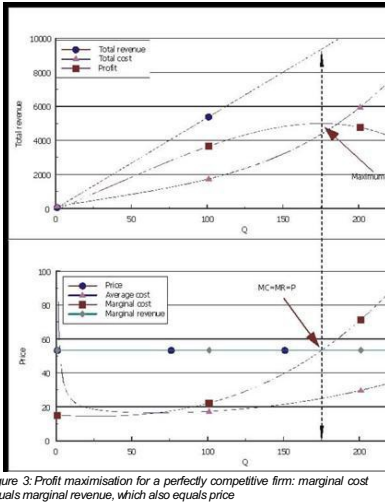


Figure 3: Profit maximisation for a perfectly competitive firm: marginal cost equals marginal revenue, which also equals price

The amount a monopolist will supply depends both on the firm's marginal cost function, and the market's demand function. Since both are needed to determine supply, and since many different demand curves can be drawn through the same point – each with a different slope and therefore different marginal revenue implications – it is impossible to derive a curve which shows how much a monopolist will supply at each price level (all you can do is consider specific examples of hypothetical demand curves, as I did to generate Table 1).

However, for the perfectly competitive firm, since price equals marginal revenue, the amount the firm will produce corresponds in every case to its marginal cost curve. The supply curve of a single firm in a perfectly competitive market is thus its marginal cost curve. Figure 4 illustrates this alleged difference between perfectly competitive firms and a monopoly.

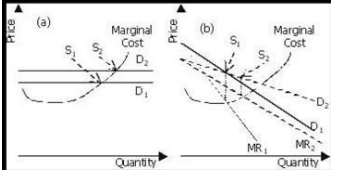


Figure 4: A supply curve can be derived for a competitive firm, but not for a monopoly

The perfectly competitive firm, in Figure 4(a), produces on its marginal cost curve no matter what the level of demand. The monopoly's level of output, however (shown in Figure 4(b)), depends on both the marginal cost curve and the marginal revenue curve. Even if two demand curves pass through the same point (S1), the amount produced for the two demand curves will differ because their marginal revenue curves differ (S1 and S2 respectively in Figure 4(b)). Supply curves therefore can't be drawn for a monopoly.

The supply curve for a perfectly competitive industry is constructed simply by adding up the amounts that each firm is willing to supply at a given price. This amounts to summing up their marginal cost curves, so that the supply curve for the industry represents the marginal cost of producing output. Since demand equals supply in equilibrium, the marginal benefit for the last unit consumed equals its marginal cost of production, and social utility is maximised. This results in both a higher level of output and a lower price than would occur if the industry were a monopoly, as Figure 5 illustrates.

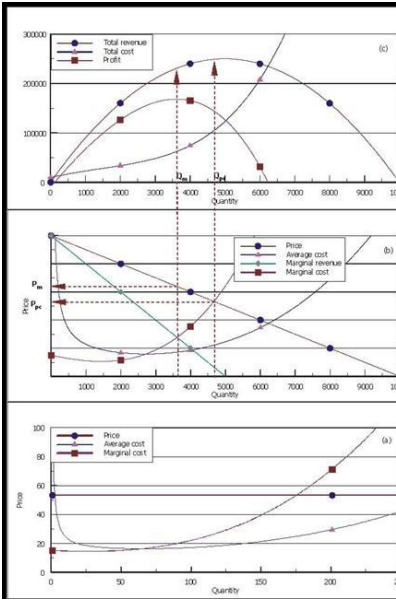


Figure 5: A supply curve can be derived for a competitive firm, but not for a monopoly

Figure 5(a) (the bottom diagram in Figure 5) shows a single perfectly

competitive firm, pricing where marginal cost, marginal revenue and price are all equal. Figure 5(b) shows the monopoly firm's marginal cost curve, as well as the combined supply curves of all perfectly competitive firms. Figure 5(c) shows the aggregate situation for both industry structures in terms of total cost, total revenue, and total profit.

The competitive firm, since it takes price as given, sells where marginal cost equals price. Total industry output is the sum of the output of all firms (this is added up effectively by compressing the quantity axis to yield the market situation shown in Figure 5(b) and 5(c)). The competitive market therefore produces Qpc units, at the point where marginal cost equals price, and charges price Pc, which equates the cost of production to the social benefit of output. Hence perfect competition is 'perfect'.

The monopoly firm in Figure 5(b) produces at Qm, the point where marginal revenue equals marginal cost, and charges price Pm. This level of production means that price exceeds marginal cost, and hence social welfare is restricted. This completes the main economic argument against monopolies and in favour of competitive industries.

Checking our sums

This argument normally convinces economics students, and it explains much of the hostility economists in general have towards monopolies, or any market in which firms have some market power by virtue of their size. This 'social radicalism' is unusual for a profession which is normally perceived as socially conservative. It is also curiously at odds with the real world, where it's fairly obvious that industries have a clear tendency to end up being dominated by a few large firms. But, economists would argue, their opposition to large firms, and their allegedly uncharacteristic radicalism on this issue, is based on sound analysis. But is it? Let's check, after first seeing how moving from a monopoly to a perfectly competitive industry would benefit society in my example.

Table 4 'adds up' the costs and revenues of all the competitive firms, to produce a hypothetical 'perfectly competitive monopoly' which behaves as the theory says the sum of all perfectly competitive firms behave, and produces where marginal cost equals price. Note that the output of this mythical firm is higher than the actual monopoly's – roughly 4670 units, versus 3615 – and its price is lower – roughly $53.28 per unit, versus $63.86 for the monopoly.

Quantity	Price	Total Revenue	Total Cost	Profit	Av. Cost
1	99.99	99.99	10,015.00	-9,915.01	10,015.0
9	99.91	899.19	10,134.80	-9,235.61	1,126.0
10	99.90	999.00	10,149.75	-9,150.75	1,014.9
100	99.00	9,900.00	11,475.26	-1,575.26	114.75
1000	90.00	90,000.00	22,861.11	67,138.89	22.86
3000	70.00	210,000.00	48,250.00	161,750.00	16.08
3001	69.99	210,039.99	48,268.76	161,771.23	16.08
3614	63.86	230,790.04	62,312.52	168,477.52	17.24
3615	63.85	230,817.75	62,340.24	168,477.51	17.24
4670	53.30	248,911.00	103,837.21	145,073.79	22.23
4671	53.29	248,917.59	103,890.50	145,027.09	22.24
4672	53.28	248,924.16	103,943.82	144,980.34	22.25
4673	53.27	248,930.71	103,997.17	144,933.54	22.25

Table 4: Cost & revenue for a 'perfectly competitive' industry identical in scale to hypothetical monopoly

Economists therefore prefer a competitive industry to a monopoly on three grounds:

- The competitive industry produces where marginal cost equals price, thus maximising social welfare;
- It produces a higher level of output than a monopoly, and
- It sells this higher output at a lower price.

There is one more reason why economists prefer perfect competition to monopoly, which Figure 6 illustrates: perfect competition is the only market structure in which price and quantity are set by the intersection of the supply curve and the demand curve (as shown in Figure 6(a)). In a monopoly, price is set by the intersection of marginal cost and marginal revenue, and in place of the simple and uncluttered two-line drawing of the standard supply and demand model, the monopoly situation requires three lines: marginal cost, marginal revenue, and demand (Figure 6(b)).

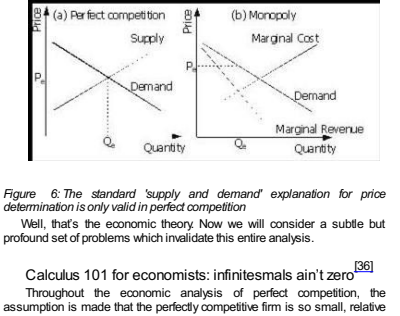


Figure 6: The standard 'supply and demand' explanation for price determination is only valid in perfect competition

Well, that's the economic theory. Now we will consider a subtle but profound set of problems which invalidate this entire analysis.

Calculus 101 for economists: infinitesmals ain't zero[36]

Throughout the economic analysis of perfect competition, the assumption is made that the perfectly competitive firm is so small, relative to the overall market, that its impact on the market can be treated as zero. As I intimated earlier in this chapter, this kind of logic is OK when you are dealing with local approximations – such as whether you can regard the ground on which you stand as either flat or curved – but it will not do when those local approximations are aggregated together. When we insist that infinitesmally small amounts are not in fact zero, the apparently watertight logic behind the comparison of monopoly and perfect competition falls apart.

Too small to matter?

An essential part of the argument for perfect competition is that each firm is so small that it can't affect the market price – which it therefore takes as given. Consequently the demand curve, as perceived by each firm, is effectively horizontal at the market price. The firms are also so small that they do not react to any changes in behaviour by other firms: in the language of economic theory, their 'conjectural variation' – how much all other firms change their output in response to a change in output by one firm – is zero.

These two assumptions are alleged to mean that the slope of the individual firm's demand curve is zero: both the firm's price and the market price do not change when a single firm changes its output. However, they also mean that, if a single firm increases its output by one unit, then total industry output should also increase by one unit – since other firms won't react to the change in output by a single firm.

However, there is a problem: these two assumptions are inconsistent.

If the market demand curve is downward sloping, then an increase in total market output must mean a fall in the market price – regardless of how small a fall it might be. Since the theory assumes that other firms don't react to an increase in production by one firm, total market output must increase. Since the market demand curve is downward sloping, and supply has increased – the supply curve has shifted outwards – market price must fall.

Therefore market price does change because of the actions of a single firm. The only way market price could not react would be if all other firms reduced their output by as much as the single firm increased it: then the market supply curve would not shift, and the price would remain constant. But the theory assumes that firms don't react to each other's behaviour.

So the market price will be affected by the actions of a single firm, in which case the demand curve facing a single firm will be downward sloping – however slight the slope may be.

Putting this critique another way, the economic argument is that if you break a large downward sloping line (the market demand curve) into lots of very small lines (the demand curves perceived by each firm), then you will have a huge number of perfectly flat lines. Then if you add all these perfectly flat lines together again, you will get one downward sloping line.

This is mathematically impossible. If you add up a huge number of flat lines, you will get one very long flat line. If you break one downward sloping line into many small lines, you will have many downward sloping lines. The economic concept of perfect competition is based on a mathematical error of confusing a very small quantity with zero.

The market matters: marginal revenue is marginal revenue

A second problem with this economic model is the nature of the marginal revenue function. Economists unconsciously reason as if the marginal revenue curve at the market level is a function of the number of firms that produce the industry's output: it exists if there is only one firm, but if there are a large number of firms, it disappears. They then show that a monopoly sets its price where marginal revenue equals marginal cost, which is consistent with their theory. However, they show a competitive industry setting price where the supply curve and the demand curve intersect, with no pesky marginal revenue curve getting in the way.

Unfortunately, marginal revenue exists independently of the number of firms in the industry. If the market demand curve is downward sloping, then so is the market marginal revenue curve. If a competitive industry results in output being set by the intersection of the demand curve and the supply curve, then at the collective level this means that a competitive industry produces where marginal cost exceeds marginal revenue. Rather than maximising profits, as economists argue firms do, the additional output – that produced past the point where marginal revenue equals marginal cost – must be produced at a loss. This paradox means that the individual firm and the market level aspects of the model of perfect competition are inconsistent.

Creative accounting

For the assertion that perfect competition results in a higher level of output at a lower price than monopoly to be correct, then in the aggregate, the individually rational profit-maximising behaviour of perfectly competitive firms must lead to a collectively irrational outcome.

According to the theory, the monopoly firm produces only to the point at which marginal cost equals marginal revenue, because this output maximises its profit. Each perfectly competitive firm likewise produces to a point at which its marginal cost equals its marginal revenue, and for the same reason – because this level of output maximises its profit.

But at the market level, they produce to a point at which the collective marginal cost exceeds marginal revenue.

Take another look at Figure 5. Notice that (b) and (c) are consistent for monopoly: marginal cost equals marginal revenue in both frames. However, for the perfectly competitive industry, (a), (b) and (c) are not consistent. At (a), the level of the firm, marginal cost equals marginal revenue equals price – under the assumptions of the model – and the individual firm produces where its profit is maximised. But at (b) and (c), the level of the industry, these equalities no longer apply. The perfectly competitive industry produces where marginal cost equals price but exceeds marginal revenue.

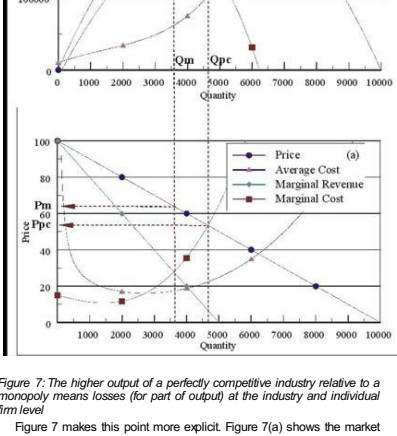


Figure 7: The higher output of a perfectly competitive industry relative to a monopoly means losses (for part of output) at the industry and individual firm level

Figure 7 makes this point more explicit. Figure 7(a) shows the market situation for both a monopoly and a competitive industry. The monopoly sets price where marginal revenue equals marginal cost, while the competitive industry sets price where the supply curve (which is the sum of all the

individual firm marginal cost curves) intersects the demand curve. Figure 7(b) shows what this means in terms of profit for the monopoly versus aggregate profits for the competitive firms. The monopoly produces output Qm, sells it at the average price Pm, and makes profit M; the perfectly competitive industry produces the larger output Qpc, sells it at the lower price Ppc, and makes the lower profit PC.

The monopoly has produced where profit is maximised, but the competitive firms have produced beyond this point, so that the industry's output past the point of monopoly output has been produced at a loss.

Where did this loss come from? It certainly can't be seen in the standard graph economists draw for perfect competition (see Figure 5.4(a)), which shows the individual competitive firm making profits all the way out to the last item produced.

Instead, this 'loss out of nowhere' is hidden in the detail that economists lose by treating infinitesimally small quantities as zeros. If perfectly competitive firms were to produce where marginal cost equals price, then they would be producing part of their output past the point at which marginal revenue equals marginal cost. They would therefore make a loss on these additional units of output.

As I argued above, the demand curve for a single firm cannot be horizontal – it must slope downwards, because if it doesn't, then the market demand curve has to be horizontal. Therefore, marginal revenue will be less than price for the individual firm. However, by arguing that an infinitesimal segment of the market demand is effectively horizontal, economists have treated this loss as zero. Summing zero losses over all firms means zero losses in the aggregate. But this is not consistent with their vision of the output and price levels of the perfectly competitive industry.

The higher level of output must mean losses are incurred by the industry, relative to the profit maximising level chosen by the monopoly. Losses at the market level must mean losses at the individual firm level – yet these are presumed to be zero by economic analysis, because it erroneously assumes that the perfectly competitive firm faces a horizontal demand curve.

Perfect competition equals monopoly

The above critique raises an interesting question: what will the price and output of a perfectly competitive industry be, if we drop the invalid assumption that the output of a single firm has no effect on the market price? The answer is: the price and output levels of a competitive industry will be exactly the same as for the monopolist (if the aggregate marginal cost curve of the competitive firms is identical to the marginal cost of the monopoly, which economic theory assumes it is).

Economic explanations of price setting in a competitive market normally start from the level of the market, where they show that the intersection of supply and demand sets both price and quantity (see Figure 6(a)). They then argue that the price set by this intersection of supply and demand is taken as given by each competitive firm, so that the supply curve for the individual firm is its marginal cost curve. Then they notionally add all these marginal cost curves up, to get the supply curve for the industry – and its point of intersection with the demand curve determines the market price.

But there is a 'chicken and egg' problem here. Which comes first – price being set by the intersection of supply and demand, or individual firms equating marginal cost to price? And why should a level of output which involves making a loss on part of output (the part past where market marginal revenue equals marginal cost) determine where each individual firm perceives price as being set?

Economists have been bewitched by their own totem. They draw a downward-sloping market demand curve, and an upward-sloping supply curve, and assume that price and quantity must be set by the intersection of the two curves. But the 'supply curve' is really only the aggregate of the marginal cost curves of all the competitive firms. It isn't a supply curve unless they can prove that, whatever the market demand curve looks like, the industry will supply the quantity given by the intersection of the demand curve and the aggregate marginal cost curve.

This isn't the case in their own model of monopoly, where economists draw the situation shown in Figure 8. The intersection of marginal cost and marginal revenue determines the quantity produced, while the price charged is set by the price the demand curve gives for that quantity: price and quantity are not determined by the intersection of the demand curve and the marginal cost curve.

Economists claim that price and quantity are set by the intersection of the demand curve and the aggregate marginal cost curve in the case of perfect competition, but their 'proof' relies on the erroneous proposition that the demand curve perceived by each individual firm is, you guessed it, horizontal.

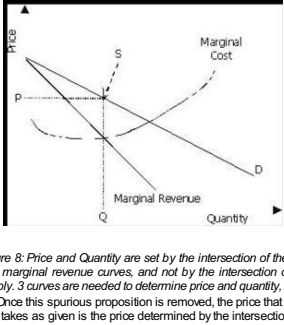


Figure 8: Price and Quantity are set by the intersection of the marginal cost and marginal revenue curves, and not by the intersection of demand with supply. 3 curves are needed to determine price and quantity, not just 2.

Once this spurious proposition is removed, the price that the competitive firm takes as given is the price determined by the intersection of the market demand curve and the aggregate marginal cost curve – which is precisely the same price as a monopoly would charge. To argue otherwise is to argue for either irrational behaviour at the level of the individual firm – so that part of output is produced at a loss – or that, somehow, individually rational behaviour (maximising profit) leads to collectively irrational behaviour – so that profit-maximising behaviour by each individual firm leads to the industry somehow producing part of its output at a loss. However, the essence of the economic vision is that individually rational behaviour leads to collectively rational behaviour.

Therefore, the price that the perfectly competitive firm will take as given when it adjusts its output is not a market price set by equating price to marginal cost, but a market price set by equating marginal revenue to marginal cost. The quantity produced at this price will be equivalent, when summed, to the output of a single monopolist. On the grounds of properly amended economic theory, monopoly and perfect competition are identical.

Two somewhat more technical arguments as to why the theory of perfect competition is unsound show that:

- setting marginal cost equal to price does not maximise profits for a competitive firm; and
- marginal cost has to be constant for a definitive comparison of perfect competition and monopoly.

These arguments are on the web at More/Size. The mathematics behind these arguments is located at Maths/Size.

Getting Technical

The analysis above was all I had time for when first writing this book—I hadn't expected to develop any new critiques of neoclassical economics, and developing this critique in full detail was a task that I thought might take years.

Those years have elapsed, and the argument has gone well beyond those explained here—though still based on the fundamental insight that "infinitesimals ain't zeros". My published papers on this topic involve complicated mathematics, but the simplest illustration of why the neoclassical theory of the firm is nonsense is a computer program that pits what I call "instrumentally rational profit-maximisers" against neoclassical theory.

These instrumentally rational profit maximising firms follow the simplest possible profit-maximising strategy:

- Choose an initial output level;
- Change output by some randomly chosen amount (which can be either positive or negative);
- If profit increases, keep on changing output by the same magnitude and in the same direction;
- If profit falls, change output by the same magnitude but in the opposite direction.

In effect, this simulation takes Milton Friedman's "billiard players" analogy (see Chapter 7, *There is madness in their method*) seriously. In defending the theory of competition against the attacks of empirical researchers like Gardiner Means, who found that firms did not in fact calculate marginal cost and marginal revenue when working out production levels, Friedman argued that this was not a problem. While firms did not explicitly work out the calculus of profit maximisation, Friedman argued that unless the rules of thumb the adopted reached the same result, they would fail to maximise profits and would therefore be eliminated by the process of competition.

He gave the analogy of expert billiard players, who obviously do not work out their shots using Newton's Laws, but whose calculations must conform

to those laws if they were to "Pot Black" with the compelling regularity that such champions display:

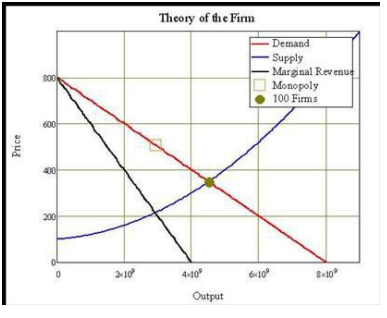
> Consider the problem of predicting the shots made by an expert billiard player. It seems not at all unreasonable that excellent predictions would be yielded by the hypothesis that the billiard player made his shots as if he knew the complicated mathematical formulas that would give the optimum directions of travel... Our confidence in this hypothesis is not based on the belief that billiard players, even expert ones, can or do go through the process described; it derives rather from the belief that, unless in some way or other they were capable of reaching essentially the same result, they would not in fact be expert billiard players.

The same principle, he asserted, applied to firms:

> It is only a short step from these examples to the economic hypothesis that ... individual firms behave as if ... they knew the relevant cost and demand functions, calculated marginal cost and marginal revenue from all actions open to them, and pushed each line of action to the point at which the relevant marginal cost and marginal revenue were equal.

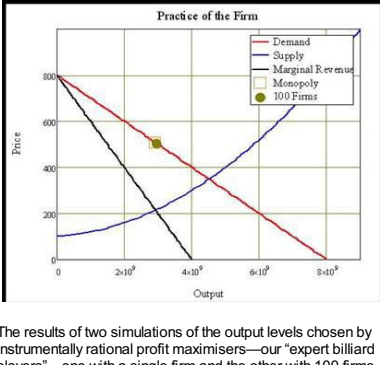
There is intrinsically nothing wrong with this argument—except, as it happens, for Friedman's confidence that the calculus undertaken by neoclassical economists deserves the same exalted status as Newton's.

My program defines a downward-sloping market demand curve, firms with upward-sloping marginal cost curves, and a market consisting of a set number of firms from one (for a monopoly) to any desired maximum number.[37] It then simulates a market with a specified number of firms, ranging from 1 to the maximum, with each firm adjusting its output 1,000 times in search of the output level that maximises its profits. Neoclassical theory predicts that the amount produced will be a function of the number of firms in the industry, as well as the supply and demand curves. The market demand and aggregate marginal cost curves are shown in Figure 9, along with the predictions the theory



makes for output levels (these are given by equating marginal cost to marginal revenue for each firm)

Figure 9: Artificial market supply and demand curves and predicted output levels



The results of two simulations of the output levels chosen by instrumentally rational profit maximisers—our "expert billiard players"—one with a single firm and the other with 100 firms, are shown in Figure 10:

Figure 10: Artificial market supply and demand curves and actual output levels

Clearly, something is wrong: contrary to the predictions of the theory, the 100 firm industry produces almost exactly the same output level as the monopoly—and the "competitive price" is therefore almost the same as the "monopoly price", since this is determined by the demand curve which is identical in both cases.

It turns out that the theory is wrong—because equating marginal cost to marginal revenue doesn't maximise profits for the individual firm, as already explained in this chapter. Only in the case of a monopoly does neoclassical theory predict the actual output level chosen; for any more firms in an industry, the output levels these instrumentally rational profit maximisers choose are a lot lower than neoclassical theory predicts, and the profit levels they earn are a lot higher (see Figure 11).

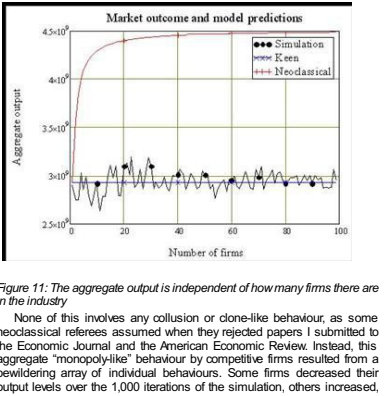


Figure 11: The aggregate output is independent of how many firms there are in the industry

None of this involves any collusion or clone-like behaviour, as some neoclassical referees assumed when they rejected papers I submitted to the Economic Journal and the American Economic Review. Instead, this aggregate "monopoly-like" behaviour by competitive firms resulted from a bewildering array of individual behaviours. Some firms decreased their output levels over the 1,000 iterations of the simulation, others increased, and many changed direction many times, with pronounced short-term trends up or down (see Figure 12). But in the aggregate, the quantity produced diverged away from the level neoclassical theory predicts to the one I worked out.

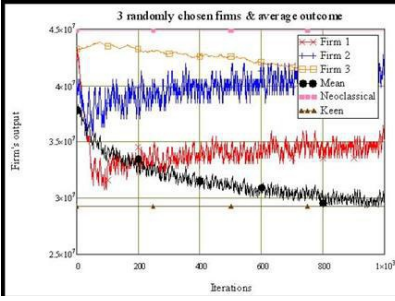


Figure 12: The behaviour of 3 randomly chosen firms in a 100 firm industry, and the average of the 100 firms

The profit levels that resulted were equally varied—though much higher in most cases than would have resulted if the firms had hired a neoclassical manager (see Figure 13).

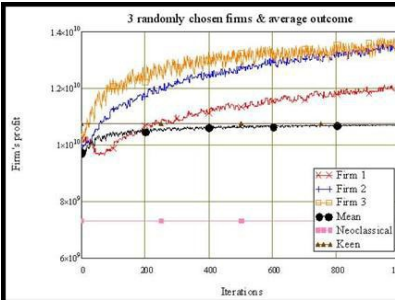


Figure 13: Profit levels for the same three firms, and the industry average

The reason why neoclassical theory is so wrong about the level of output that maximises profits for an individual firm is that its assumes that only a firm's own output level and the market price determine its profit. It then treats the market price as independent of what the individual firm does—and recommends a level of output that is almost twice the profit-maximising level.

A market is by definition a place where the actions of your competitors affect the market price, and this also has to affect the individual firm's profits. The correct calculus to work out the profit-maximising level for a single firm takes this into account. This concludes that the profit-maximising output level is not where marginal cost equals marginal revenue, but where the gap between marginal revenue and marginal cost is slightly less than the gap between price and marginal cost (see Figure 14).

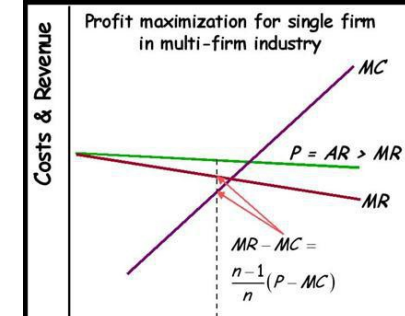


Figure 14: The true profit-maximsing position for an individual firm

A full explanation of the issues touched on here will have to await a second edition of *Debunking Economics*. Those who have the interest and to some extent capacity to cope with mathematical reasoning should turn to the papers by myself and Russell Standish noted in the references.

Returns to scale and the durability of perfect competition

To date I have accepted the assumption that a monopoly has no scale advantages over a perfectly competitive firm, so that it is possible to sum the cost functions of numerous small firms and come up with aggregate costs similar to those of a large firm (if we ignore problems discussed on the web at More/Size).

In general, this assumption of scale-invariant costs will be invalid. If we are simply considering the costs of producing a homogenous product, then it is likely that a very large firm will have scale advantages over a very small one. In the vernacular of economics, large firms benefit from returns to scale.

Returns to scale occur when the cost of production rises less rapidly than the output as the scale of production increases. A simple example is shown in Figure 15. Farms need to be separated from each other by fences. The amount of fencing required depends on the perimeter of the farm. If we consider a square block of land, fencing will depend on the length of the four sides of the square. Cost is thus the cost of fencing per mile, times four times the length of a side. But the area enclosed by the fence depends on the length of a side squared. The output of a farm is related to its area, so that output is a function of the length of a side squared. Doubling the perimeter of a farm thus doubles its fencing costs, but increases its output fourfold. As a result, large farms have a scale advantage over smaller farms. A farm with a square mile of land requires four miles of perimeter fencing to each square mile, while a farm with four square miles of land requires eight miles of perimeter fencing – or just two miles of perimeter fencing to each square mile of land.

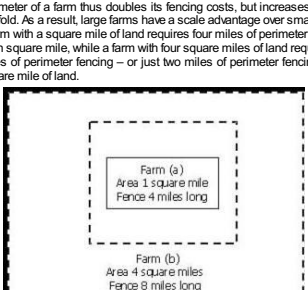


Figure 15: Double the size, double the costs, but four times the output

The same concept applies in numerous ways. For a substantial range of output, a large blast furnace will be more cost effective than a smaller one, a

large ship than a smaller one, a large car factory than a smaller one.

If large firms have cost advantages over small ones, then given open competition, the large firms will drive the small ones out of business (though marketing and debt problems will limit the process, as Sraffa notes). Hence increasing returns to scale mean that the perfectly competitive market is unstable: it will, in time, break down to a situation of either oligopoly (several large firms) or monopoly (one large firm).

Economists have been well aware of this dilemma since Marshall at the end of the 19th century, and the fiction that has been invented to cope with it is the concept of the long run average cost curve.[38] The curve is 'u-shaped', which asserts that there is some ideal scale of output at which the cost of production is minimised. In the long run, all inputs can be varied, so this shape is supposed to represent increasing returns to scale up to the point of minimum cost, beyond which decreasing returns to scale start to occur, so that the cost of production rises.

A competitive industry is supposed to converge to this ideal scale of output over time, in which case its many extremely big firms are safe from the predations of any much larger firm, since such a competitor would necessarily have higher costs.

This defence is specious on several counts.

Firstly, the question of whether perfect competition can exist in a particular industry becomes an empirical one: what is the ideal scale of output, and how many firms could then occupy a particular industry at a particular time?

For some industries, the answer might well be 'very many' – the ever popular wheat farm comes to mind. However for some other industries, the answer might well be 'very few'. It seems, for example, that the worldwide market for large intercontinental passenger aeroplanes can support at most three firms.

The argument that, 'in the long run', this industry could be perfectly competitive because it could grow big enough to support hundreds or thousands of competitors is ludicrous. By the time the world was large enough to support hundreds of Boeings, it is highly likely that some entirely different form of transport would have superseded the aeroplane.

Secondly, the long run supply curve is actually constructed under the assumption of constant technology: in other words, it is not really a concept in time at all. The scale economies are supposedly there all the time, ready to be exploited.

If so, then unless an industry is already big enough to support the enormous number of firms surmised by the model of perfect competition – all operating at the ideal scale – large firms can immediately out-compete small firms. In other words, the only way competitive firms can survive is if the industry is already so large that it can support an enormous number of firms of the ideal scale.

The theoretical response of economists to this dilemma has been to presume constant returns to scale. With constant returns, 'size does not matter': a small firm will be just as cost efficient as a large one.

Unfortunately, size does matter. Economies of scale are an important part of the reason that most industries are dominated by a small number of very large firms. We do need an adequate analysis of how such an industry functions, but economics does not provide it.[39]

So What?

Economics has championed the notion that the best guarantee of social welfare is competition, and perfect competition has always been its ideal. The critiques in this chapter show that economic theory has no grounds whatsoever for preferring perfect competition over monopoly. Both fail the economist's test of welfare, that marginal cost should be equated to price.

Worse, the goal of setting marginal cost equal to price is as elusive and unattainable as the Holy Grail. For this to apply at the market level, part of the output of firms must be produced at a loss. The social welfare ideal thus requires individual irrationality. This would not be a problem for some schools of economics, but it is for the neoclassical school, which has always argued that the pursuit of individual self-interest would lead to the best, most rational outcome for all of society.

Economics can no therefore longer wave its preferred flag, but must instead choose between the three shown in Figures 8 and 10. Either the demand curve can be drawn downward sloping (if we ignore the consequences of the Sonnenshein-Mantel-Debreu conditions), but supply can only be shown as a point determined by intersection of the marginal cost and marginal revenue curves, as in Figure 8. The basic 'supply and demand' diagram must have three lines, not just two.

Or, worst of all, if the theory wants to be able to compare perfectly competitive firms to monopolies, then – for reasons outlined on the web at More/Size – the marginal cost curve must be drawn horizontally. In this case, the model of monopoly can still reach a determinate outcome, but the model of perfect competition collapses, as shown in Figure 10.

Perfect competition is thus decidedly less perfect than economists have believed, and not only does it fail to provide a higher welfare level than monopoly, it may also be unstable and liable to collapse into monopoly over time – as Sraffa pointed out in 1926.

The economic bias in favour of small firms and against monopolies is, on the evidence of this chapter, almost totally misplaced. As it happens, a slightly different approach to the theory of competition (known as the theory of contestability) has softened this opposition in recent decades – since it argues that what matters is not the number of actual competitors now, but the threat of new competitors if monopolies set price too high relative to marginal cost. But the residual bias against monopolies remains, and for no good reason. As I noted at the beginning of this chapter, there are plenty of good reasons to be wary of monopolies and to wish to modify their behaviour, but economic theory does not provide any of them.

The main danger of economic theory in this area is that it will encourage the dismemberment of existing monopolies for no good reason, or the use of competitive approaches to the provision of public services when a concerted monopoly approach would be superior. Telecommunications provides examples of both these problems.

Meanwhile, back in the real world

America's anti-trust break-up of 'Ma Bell' into the 'Baby Bells' had quite a range of effects, and economists like to dwell on the increase in competition it inspired and the alleged benefits of this to consumers. However, the unavoidably regional nature of the break-up made long-distance communications in the USA significantly more complicated than it had been, and many engineers allege that it resulted in a drop in technical standards. The most apparent remnant of this problem today is the poor state of the mobile telephone network in the USA, and the consequent limited penetration of mobile telephony there compared to less obvious centres of telecommunications excellence – such as Finland.

Australia's 'competitive' roll-out of optical cable, in which two suppliers competed to provide physical cable links to households and firms, led to the richer and densely populated parts of the country having two optical cables running past every house, while poorer or less densely populated parts had none. The disparity has widened as many new suppliers provide optical links in and between major cities, ignoring the unprofitable country towns, homesteads, and even wealthy 'large estate' regions of major cities.

New forms of technology have ameliorated the damage, by providing other means by which those not connected to optical fibre can have high-speed telecommunications. But the end result of following a competitive approach has been that far too much has been spent on providing Australia's telecommunications network, and telephone and Internet access prices have been kept artificially high to finance this waste.

The proper manner in which the issue of 'monopoly versus competition' should be approached is a case by case basis, where the merits of one form or the other of organising production can be dispassionately considered. The instinctive economic bias against monopoly should be replaced by something rather more intelligent.

Not a product endorsement

The fact that there is no substance to the economic critique of monopolies does not mean that monopolies are good. There are plenty of reasons why monopolies can be 'a bad thing': they can bully and stifle potential competitors, set higher markups than more competitive markets, and use their monopoly power to exploit other markets. All these practices occurred in the 19th century with the 'robber barons', and these practices explain why America has its system of anti-trust laws.

Unfortunately, the economic argument against monopolies has redirected attention away from these qualitative nuances to the simple issue of whether a monopoly exists – defined by the proportion of the market dominated by one firm. This simplistic criterion treats all monopolies as tainted, whereas the old anti-trust philosophy attacked monopolies for particular types of behaviour.

The economic argument can therefore lead to monopolies being broken up where there were no signs of such behaviour, or where there were extremely good reasons why a single firm was far more efficient than a number of competing firms. Far from being a guide about how to achieve greater economic efficiency, economics may encourage us to dismantle effective companies and replace them with inefficient ones.

A theory without substance

There is no point in trying to construct a diagram to represent what remains of the standard microeconomic supply and demand analysis after the previous three substantive chapters, because essentially nothing remains. The demand curve can take any shape at all – except, of course, the smooth, downward-sloping shape which economists normally draw. Diminishing marginal productivity is unlikely to apply in practice, and, to the extent that the concept of a supply curve is meaningful, it is probably

horizontal or downward sloping.

Therefore, if demand and supply curves can be drawn at all, they will be jagged rather than smooth, smudged rather than definitively sharp, and they will intersect in numerous places, none of which can be said to be superior to any other. Economists might wish to believe that 'price is set by supply and demand', but their own theories support no such straightforward conclusion.

Yet still they teach the standard mantra to their students, and still they apply the same simplistic logic to many other areas of economics.

In the chapters to come, we will temporarily 'forget' the criticisms of these fundamental building blocks of economic logic, and examine the validity of economic theory as it is applied to specific issues. As you shall see, even if we allow, for the sake of argument, that demand falls smoothly as price rises, that production is subject to diminishing marginal returns, and that demand and supply set prices, the economic theories of the distribution of income, the behaviour of the macro-economy, and the role of finance, are all intellectually unsound.

5. To each according to his contribution

Why productivity doesn't determine wages

One of the most striking aspects of the late 20th century was the increase in the gap between the poorest worker and the richest.[40] While many bemoaned this increase in inequality, economists counselled that the growing gap merely reflected the rising productivity of the highly paid.

The basis for this advice is the proposition that a person's income is determined by her contribution to production – or more precisely, by the marginal productivity of the 'factor of production' to which she contributes. Wages and profits – or 'factor incomes' as economists prefer to call them – reflect respectively the marginal product of labour and of capital. The argument that highly paid workers – managers of major corporations, stock and money market traders, financial commentators, etc. – deserve the high wages they receive compared to the less highly paid – nuclear physicists, rocket scientists, university professors, school teachers, social workers, nurses, factory workers, etc. – is simply an extension of this argument to cover subgroups of workers. Members of the former group, we are told, are simply more productive than members of the latter, hence their higher salaries.

I'll defer discussion of the proposition that profits reflect the marginal productivity of capital until the next chapter. Here we'll consider the argument that wages equal the marginal product of labour.

Once again, the argument relies heavily on concepts we have already dismissed: that productivity per worker falls as more workers are hired; that demand curves are necessarily downward sloping; that price measures marginal benefit to society; and that individual supply curves slope upwards and can easily be aggregated. Even allowing these invalid assumptions, the economic analysis of the labour market is still flawed.

The kernel

Economists prefer to treat everything, including labour, as a simple commodity, subject to the same 'laws of supply and demand' as the simple apple. Yet their own analysis of labour shows that it is fundamentally different. In all other markets, demand decisions are made by consumers and supply decisions by producers. But in the labour market, supply decisions are made by consumers (households supplying labour), whereas labour demand decisions are made by producers (firms hiring labour). Thus the conventional economic analysis of markets, which is suspect enough on its own terms, is highly unlikely to apply in this most crucial of markets. As a result, wages are highly unlikely to reflect workers' contributions to production, as economists argue.

The roadmap

In this chapter, I outline the economic analysis of labour supply, and the normal economic argument in favour of letting the market decide both wages and the level of employment.

I show that irregularities in the supply of labour – when compared to a normal commodity – are easily derived from this analysis, yet economists unjustifiably assume that labour supply will be an upward-sloping function of the wage. However, these labour market irregularities can make the supply of labour 'backward-bending', so that reducing wages could actually cause the supply of labour to rise rather than fall.

Though economists normally oppose unions, there are economic arguments in favour of a cartel when sellers (such as workers selling their labour) face a buyer with market power. The opposition economists normally present to unions, to interventionist labour market policies, and to attempts to reduce income inequality, are thus shown to be unjustified, even on the grounds of standard economic logic.

Labour demand and supply: an inverted commodity

The economic theory that a person's income reflects his contribution to society relies on being able to treat labour as no different from other commodities, so that a higher wage is needed to elicit a higher supply of labour, and reducing the wage will reduce supply. In fact, economic theory supports no such conclusion. Even economists can't escape the fact that, as commodities go, labour is something out of the ordinary.

The demand for ordinary commodities is determined by consumer incomes and tastes, while supply is determined by the costs of production. However, unlike other commodities, no-one actually 'consumes' labour: instead, firms hire workers so that they can produce other commodities for sale. Secondly, unlike all other commodities, labour is not produced for profit – there are no 'labour factories' turning out workers according to demand, and labour supply certainly can't be said to be subject to the law of diminishing returns (whatever parents might think!).

These two peculiarities mean that, in an inversion of the usual situation, the demand for labour is determined by producers, while the supply of labour is determined by consumers. Demand reflects firms' decisions to hire workers to produce output for sale; supply reflects workers' decisions about how long to work, on the basis of their preferences for income on the one hand and leisure time on the other.

If economists are to argue that the labour market is to behave like all other markets, then these peculiarities must not complicate the usual totemic duet of a downward-sloping demand curve and upward-sloping supply curve. Unfortunately for economists, they do.

Marginal workers

According to economic theory, a firm's labour-hiring decision is determined simply by the impact that each additional worker has on the firm's bottom line. If hiring an additional worker will add to the firm's profit, she is hired; if not, the firm stops hiring.

With a perfectly competitive labour market, the firm can hire as many workers as it wishes to at the going wage. However, since one input (capital) is fixed in the short run, output is subject to diminishing returns: each additional worker hired adds a lesser amount to output than her predecessor. Diminishing marginal productivity therefore rules the hiring roost.

For each firm, the wage is a constant (set by the labour market in which each firm is an infinitesmally small actor). The amount each worker adds to profits, however, is variable. The firm keeps hiring workers up until the point at which the wage equals the amount for which that the last worker's additional output can be sold.

If the industry itself is perfectly competitive, the additional units can be sold without the firm having to reduce its price (yes, I know, that's been debunked two chapters ago; but let's pretend otherwise). In general, the revenue the firm gains by hiring its last employee is equal to the price for which it sells its output, multiplied by the marginal product of the last worker. The firm's demand for labour is therefore the marginal physical product of labour multiplied by the price of the output.

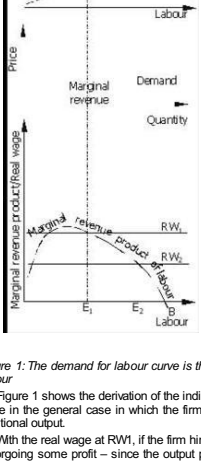


Figure 1: The demand for labour curve is the marginal revenue product of labour

Figure 1 shows the derivation of the individual firm's demand for labour curve in the general case in which the firm has to reduce its price to sell additional output.

With the real wage at RW1, if the firm hires any less than E1 workers, it is forgoing some profit – since the output produced by additional workers

exceeds the cost of hiring them. If it hires any more workers than E1, then it has cost itself money: the additional output produced will sell for less than the cost of hiring the workers. The same argument applies to the real wage level RW2, and all other wage levels. Therefore a firm's demand for labour curve is its marginal revenue product curve.[41]

A disaggregated picture of this is used to explain why some workers get much higher wages than others. They – or rather the class of workers to which they belong – have a higher marginal revenue product than more poorly paid workers. Income disparities are the product of differential contributions to society, and though sociologists may bemoan it, both the rich and the poor deserve what they get.

Aggregate demand

The demand curves for individual firms are aggregated to form this industry's demand curve for labour, which itself will be a small part of the economy-wide demand curve for labour (since workers can generate many different kinds of output). The real wage is set by the point of intersection of this aggregate demand for labour curve – labour's aggregate marginal revenue product curve – with the aggregate supply curve.

Aggregate supply, in turn, is simply the sum of the supply decisions of individual workers. According to economists, a worker's decision about how much labour to supply is made the same way she decides how much to consume.

Indifferent workers

Individual labour supply is determined by the individual's choice between work and leisure. Work is a 'bad' in Bentham's calculus: work is a 'pain' while leisure is a 'pleasure'. Therefore the pain of work must be compensated for by pleasure of the wage, to make up for the sacrifice of leisure required to earn the wage.

This choice is represented, as always, by an indifference map. Potential income is on the vertical axis, and potential leisure time on the horizontal. The indifference map represents a consumer's preferences between leisure and income, and this can take any shape consistent with the economic axioms of rational behaviour outlined in Chapter 2. The budget line represents the hourly wage rate: the higher the wage, the steeper the budget line.

This model has one peculiarity when compared to that applied to normal commodities. With standard commodities, the budget line can be drawn anywhere, so long as it reflects the relative price of the commodities in its slope, and the consumer's income in its horizontal and vertical intercepts.[42] But with labour, the horizontal axis end of the budget line is fixed at 24 hours, since that's the maximum amount of leisure anyone can have in a day. For this reason, all that the budget line can do in this model is pivot about the horizontal intercept, with the slope representing the hourly wage. The distance from zero to the 24 hour mark represents the maximum possible leisure of 24 hours a day.

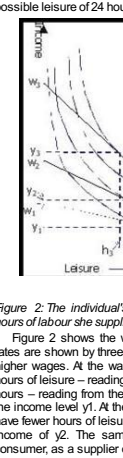


Figure 2: The individual's income-leisure trade-off determines how many hours of labour she supplies

Figure 2 shows the workings of the model. In the figure, three wage rates are shown by three budget lines, with the steeper lines representing higher wages. At the wage level w1, the consumer chooses to have h1 hours of leisure – reading from zero – and therefore chooses to work 24-h1 hours – reading from the 24 hour mark back towards zero. This results in the income level y1. At the higher wage rate w2, the consumer chooses to have fewer hours of leisure, and to work 24-h2 hours for the yet higher total income of y2. The same applies at the next higher wage rate – the consumer, as a supplier of labour, is supplying more labour as the price of

labour rises, exactly like a normal supply curve.[43]

The supply curve itself can easily be drawn by plotting the wage rate (price) on the y-axis and hours worked (quantity) on the x-axis. This is shown in Figure 3.

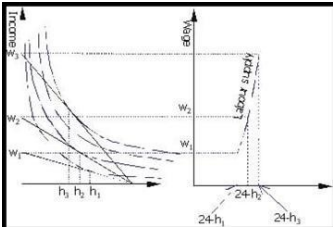


Figure 3: An upward sloping individual labour supply curve

The individual supply curve is then summed with that of all other workers to produce the market supply curve. We are back in the familiar economic territory of a downward-sloping demand curve and an upward-sloping supply curve intersecting to determine an equilibrium price: the average wage. The 'totem of microeconomics' is once again held aloft in Figure 4. This argument, which strictly speaking applies to labour in the aggregate, is extended by analogy to a disaggregated level in order to explain why some workers get much higher wages than others.

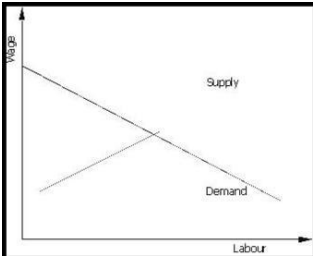


Figure 4: Supply and demand determine the equilibrium wage in the labour market

At a policy level, this model is used to emphasise the futility of minimum wage legislation, demand management policies, and any other attempts to interfere with the free working of the market mechanism in this most political of markets.

If a government attempts to improve workers' incomes by legislating a minimum wage, then this will result in unemployment, as shown in Figure 5. The equilibrium wage is We, and the equilibrium supply of labour Le. If the government instead sets the minimum wage at Wm, then workers will offer Ls hours of work at that wage rate. However, employers will only find it profitable to offer Ld hours of work at that wage. The gap between Ld and Ls represents involuntary unemployment at this artificially high wage level.

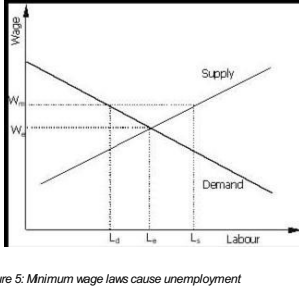


Figure 5: Minimum wage laws cause unemployment

If instead a government attempts to boost employment by demand management measures, this amounts to attempting to set the employment level at Lf, as shown in Figure 6. Demand management policies won't move the demand curve for labour, because this is the product of the marginal physical product of labour and the price for which labour's output can be sold. Demand management policies can't alter the marginal physical product of labour – this can only be done by raising the productivity of labour, which is a supply-side issue. The price of labour's output also can't be raised at this aggregate level, because we are now dealing with an economy-wide model: increasing aggregate prices merely causes inflation, without increasing the real returns to firms.

Demand management policies can only work by temporarily fooling workers into believing they are receiving a higher real wage than they actually are – as if the actual demand curve, determined by labour's marginal revenue product, had moved out. Friedman argued that the initial effect of inflation caused by government demand management policies was to fool workers into believing that the real wage is Wp, when in fact the real wage compatible with employment level Lf is Wa. Once workers realise that the higher money wage actually represents a much lower level of real wages, labour supply will adjust back down to the equilibrium level.

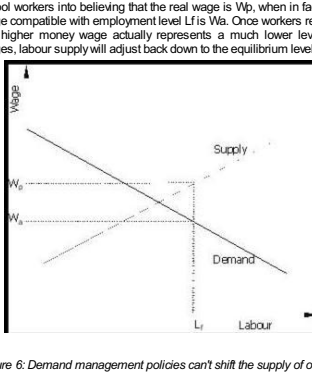


Figure 6: Demand management policies can't shift the supply of or demand for labour

The essential message is that 'you can't beat the market'. Whatever society may think is a fair wage level, or a socially desirable level of unemployment, ultimately the market will decide both income distribution and the rate of unemployment. Moreover, both these market outcomes will be fair: they will reflect individual productivity on the one hand, and the labour–leisure preferences of individuals on the other.

Problems

There are at least four serious problems with this meritocratic view of income distribution and employment determination. The supply curve for labour can 'slope backwards' – so that a fall in wages can cause an increase in the supply of labour; when workers face organised or very powerful employers, economic theory shows that workers won't get fair wages unless they also organise; Sraffa's observations about aggregation, noted in Chapter 3, indicate that it is inappropriate to apply standard supply and demand analysis to the labour market; and the basic vision of workers freely choosing between work and leisure is flawed.

Backward-bending supply curves

The upward-sloping supply curve derived in Figures 2 and 3 is not the only possible shape. It is just as easy to derive a downward-sloping supply curve – where the amount of labour supplied falls as the wage rises. This contrary case is shown in Figures 7 and 8.

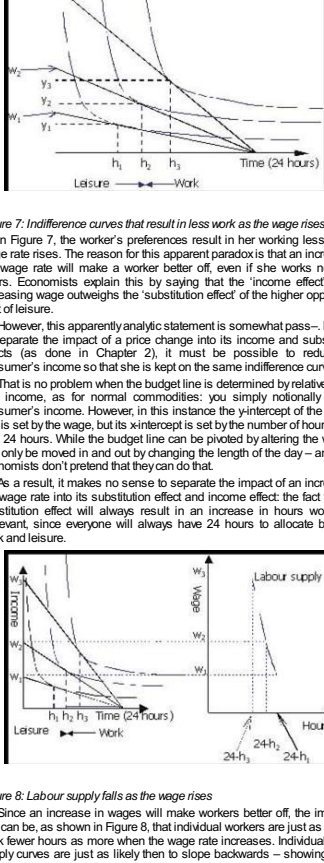


Figure 7: Indifference curves that result in less work as the wage rises

In Figure 7, the worker's preferences result in her working less as the wage rate rises. The reason for this apparent paradox is that an increase in the wage rate will make a worker better off, even if she works no more hours. Economists explain this by saying that the 'income effect' of the increasing wage outweighs the 'substitution effect' of the higher opportunity cost of leisure.

However, this apparently analytic statement is somewhat pass–. In order to separate the impact of a price change into its income and substitution effects (as done in Chapter 2), it must be possible to reduce the consumer's income so that she is kept on the same indifference curve.

That is no problem when the budget line is determined by relative prices and income, as for normal commodities: you simply notionally alter a consumer's income. However, in this instance the y-intercept of the budget line is set by the wage, but its x-intercept is set by the number of hours in the day: 24 hours. While the budget line can be pivoted by altering the wage, it can only be moved in and out by changing the length of the day – and even economists don't pretend that they can do that.

As a result, it makes no sense to separate the impact of an increase in the wage rate into its substitution effect and income effect: the fact that the substitution effect will always result in an increase in hours worked is irrelevant, since everyone will always have 24 hours to allocate between work and leisure.

Figure 8: Labour supply falls as the wage rises

Since an increase in wages will make workers better off, the impact of this can be, as shown in Figure 8, that individual workers are just as likely to work fewer hours as more when the wage rate increases. Individual labour supply curves are just as likely then to slope backwards – showing falling supply as wages rise – as they are to slope forwards.

One way to characterise the feasible shape of a labour supply curve derived from the vision of a utility-maximising worker is to consider extreme positions – very high and very low wages – as well as intermediate levels. Compared to the totemic vision of a downward-sloping supply curve, the shape so derived is perverse.

For extremely low wage levels, individual labour supply is likely to be very

high. This is because the point of tangency between a very flat budget line (signifying a very low wage) and a leisure/income indifference map is likely to be very near the point of zero leisure. An economic interpretation of this would be that, with lots of leisure time and very little income, the 'marginal rate of substitution' between income and leisure is quite high – the consumer or would-be worker is willing to sacrifice quite a bit of leisure time (of which she has a lot) for a small amount of income (of which she has very little). A more realistic interpretation would be that, if wages are just above starvation levels, you have to work a lot of hours just to stay alive.[44]

For extremely high wage levels, the point of tangency is likely to be very near the point of zero labour: why work more than a bare minimum if wages are astronomical and work is a 'bad'.[45]

Finally, as shown above, at intermediate wage levels the outcome for individual labour supply is indeterminate, so that the labour supply curve is probably relatively vertical. This generates an individual labour supply curve which resembles that shown in Figure 9.

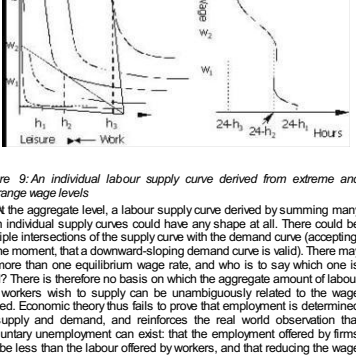


Figure 9: An individual labour supply curve derived from extreme and midrange wage levels

At the aggregate level, a labour supply curve derived by summing many such individual supply curves could have any shape at all. There could be multiple intersections of the supply curve with the demand curve (accepting, for the moment, that a downward-sloping demand curve is valid). There may be more than one equilibrium wage rate, and who is to say which one is valid? There is therefore no basis on which the aggregate amount of labour that workers wish to supply can be unambiguously related to the wage offered. Economic theory thus fails to prove that employment is determined by supply and demand, and reinforces the real world observation that involuntary unemployment can exist: that the employment offered by firms can be less than the labour offered by workers, and that reducing the wage won't necessarily reduce the gap.

This imperfection in the theory – the possibility of backward-bending labour supply curves – is normally pointed out to students of economics, but then glossed over with the assumption that, in general, labour supply curves will be upward sloping. But there is no theoretical – or empirical – justification for this assumption.

This strong assumption would be of little consequence if economists didn't derive such strong conclusions from their model of the labour market. Declarations that minimum wage legislation is ineffective and causes unemployment, or that demand management policies can't alter the rate of unemployment, are hardly insignificant pronouncements. Their truth is dependent in part on the supply curve for labour being upward sloping.

For example, if the aggregate demand and supply curves for labour look like those in Figure 10, then on the economic definition of unemployment – the gap between supply and demand for labour at a given wage – increasing the wage via legislation could make the labour market stable and decrease unemployment.

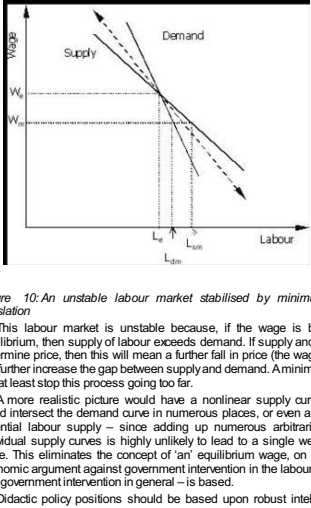


Figure 10: An unstable labour market stabilised by minimum wage legislation

This labour market is unstable because, if the wage is below the equilibrium, then supply of labour exceeds demand. If supply and demand determine price, then this will mean a further fall in price (the wage), which will further increase the gap between supply and demand. A minimum wage will at least stop this process going too far.

A more realistic picture would have a nonlinear supply curve, which could intersect the demand curve in numerous places, or even a 'cloud' of potential labour supply – since adding up numerous arbitrarily sloped individual supply curves is highly unlikely to lead to a single well-defined curve. This eliminates the concept of 'an' equilibrium wage, on which the economic argument against government intervention in the labour market – and government intervention in general – is based.

Didactic policy positions should be based upon robust intellectual or empirical foundations, rather than the flimsy substrate of mere fancy. Economists are quite prone to dismiss alternative perspectives on labour market policy on this very basis – that they lack any theoretical or empirical foundations. Yet their own policy positions on the labour market are based as much on wishful thinking as on wisdom.

Monopoly and monopsony

The conclusion that workers receive the value of their marginal contribution to output depends upon the assumption that both the product market and the labour market are perfectly competitive. The notion of perfect competition has been debunked in Chapter 4, but even if it were intellectually sound, it is clearly a dubious thing to assume for an overall economy.

If we instead accept that in practice both product and labour markets will not be perfectly competitive, then economic theory predicts that workers will not, in general, receive the value of their marginal contribution to production. In this more general case, economic theory concedes that workers' incomes are determined not only by their contribution to production, but also by the relative bargaining power of workers and employers.

Let's first consider the case in which the product market is not perfectly competitive: workers are being hired by firms who have to reduce their average selling price to increase output. In this case, the price received per unit falls as output increases. Marginal revenue is thus less than price, and the worker's marginal revenue product is the product of marginal revenue and marginal productivity. This results in the situation shown in Figure 1: the workers' marginal revenue product falls more rapidly than their marginal physical product because marginal revenue falls as output grows.

One ironic consequence of this analysis is that economic analysis can be shown to favour the existence of trade unions. Without trade unions, the labour supply will be competitive and will therefore be 'exploited', because the wage will be less than the price for which the marginal worker's output can be sold. With a trade union acting as a single seller of labour, however, the price charged for each additional worker will rise as more workers are hired. This situation – known as a monopsony or single seller – means that the marginal cost of supply lies above the supply curve.

With a monopoly seller of labour confronting non-competitive purchasers of labour, the wage is indeterminate. It will lie between the minimum set by the marginal revenue product of labour (which means that firms are exploiting workers), and the maximum set by the rising marginal cost of workers (which means that workers are exploiting firms). The final position will be determined by the relative bargaining power of the two groups, which cannot be determined by the market.

Thus while economists normally portray unions as bad because they restrict competition in the labour market, this may be a preferable situation

to leaving competitive workers to be exploited by less-than-perfectly competitive hirers of labour.

Sraffa's observations on aggregation

You will remember from Chapter 3 that Sraffa had two criticisms of economic demand and supply analysis: one for a broad definition of an industry, the other for a narrow definition. The labour market definitely is clearly a broadly defined industry, and Sraffa's first critique is therefore relevant to it.

The critique was that, with a broad definition of an industry, it is not feasible to draw independent demand and supply curves, since any change in supply will have income distributional effects which will in turn alter demand.

This is clearly the case when the supply curve refers to the entire labour force. Remember that the aggregate demand curve, in this market, is supposed to represent the aggregate marginal revenue product for labour. This in turn is a product of physical labour productivity on the one hand, and the price for which output produced by that labour is sold.

If an increase in supply requires an increase in the price of labour – if, in other words, the supply curve for labour is upward sloping – then this is clearly going to alter income distribution, the demand for commodities, and hence their prices. This means that a different 'demand curve' for labour will apply at every different point along a labour supply curve.

The picture shown in Figure 11 will then apply: multiple equilibria will exist, none of which can be said to be more fundamental than any other. It is also quite feasible that 'perverse' outcomes will apply: that, for example, a higher wage could be associated with a higher level of employment rather than a lower one (this dilemma is explored in detail in Chapter 6, in the context of the demand for capital).

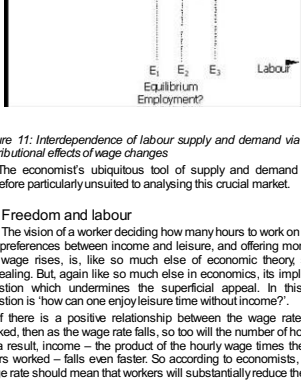


Figure 11: Interdependence of labour supply and demand via the income distributional effects of wage changes

The economist's ubiquitous tool of supply and demand analysis is therefore particularly unsuited to analysing this crucial market.

Freedom and labour

The vision of a worker deciding how many hours to work on the basis of her preferences between income and leisure, and offering more labour as the wage rises, is, like so much else of economic theory, superficially appealing. But, again like so much else in economics, its implicitly begs a question which undermines the superficial appeal. In this case, the question is 'how can one enjoy leisure time without income?'.

If there is a positive relationship between the wage rate and hours worked, then as the wage rate falls, so too will the number of hours worked. As a result, income – the product of the hourly wage times the number of hours worked – falls even faster. So according to economists, a fall in the wage rate should mean that workers will substantially reduce their incomes, and simultaneously devote more time to 'leisure activities'.

In reality, the only 'leisure activity' which one can devote more time to with less income is sleeping (just ask a homeless person). Most leisure activities are just that – active – and cost money. The only way that workers could behave as economics fantasises is if they have alternative sources of income.

This in effect is the economic vision of a worker: someone who has alternative means to generate income at her disposal, and has to be enticed by the wage to undertake wage labour for an employer over the alternative of working for herself.

For that choice to be a reality, workers need something else: capital, his/her own means of production.

Some workers are so endowed. Some farmers can be enticed into working as farm labourers if the wage is high enough, and if it's not, then they can work their own land. Some office workers have the alternative of working for a wage, or operating as independent consultants out of their home offices. Some 'wage slaves' can make the transition from employee to employer by an innovative idea, hard work, good luck, skill or good timing – or fraud.

But the majority do not have that choice – or rather don't have it to the degree that they could avoid bankruptcy or starvation by turning to self-employment. For this majority, work is not an option but – in the absence of a social security system – a necessity. Rather than smoothly choosing between work and leisure, in a completely free market system they face the choice of either working or starving. In a market economy attenuated by the welfare state, this choice is less stark, but still present.

This analysis supports Keynes's use of a horizontal supply curve for labour at the going wage.

So What?

Few issues provide better examples of the negative impact of economic theory on society than the distribution of income. Economists are forever opposing 'market interventions' which might raise the wages of the poor, while defending astronomical salary levels for top executives on the basis that if the market is willing to pay them that much, they must be worth it. In fact, the inequality which is so much a characteristic of modern society reflects power rather than justice. This is one of the many instances where unsound economic theory makes economists the champions of policies which, if anything, undermine the economic foundations of modern society.

Economics should accept that labour is unlike any other commodity, and develop an analysis suited to its peculiarities, rather than attempt to bend this most personal of markets to fit the conventional cloth of supply and demand.

Keynes did just that in the General Theory. But mainstream economics after Keynes pulled away from this innovation on the basis that Keynes's argument 'did not have good microeconomic foundations'. As this and the preceding three chapters have shown, conventional microeconomic theory itself has unsound foundations. And things get even worse when we turn our attention to problems with the other 'factor of production', capital.

Complexities

Issues that should form part of an education in economics, but which are omitted from standard courses

6. The holy war over capital

Why the productivity of capital doesn't determine profits

The economist Dharma Kumar is said to have once remarked that "Time is a device to stop everything from happening at onceand space is a device to stop everything from happening in Cambridge."

Nevertheless, a lot did happen at Cambridge during the 1960s and 1970s, where 'Cambridge' refers to both Cambridge, Massachusetts USA and Cambridge, England. The former is home to the Massachusetts Institute of Technology (better known by its initials MIT), the latter is the home of the famous University of Cambridge. MIT was the bastion for the leading true believers in economics, while the University of Cambridge housed an important group of heretics.

For twenty years, these two Cambridges waged a theoretical 'Holy War' over the foundations of neoclassical economics. The first shot was fired by the heretics, and after initial surprise the true believers responded strongly and confidently. Yet after several exchanges, the leading bishop of the true believers had conceded that the heretics were substantially correct. Summing up the conflict in 1966, Paul Samuelson observed that the heretics "merit our gratitude" for pointing out that the simple homilies of economic theory are not in general true. He concluded that

> If all this causes headaches for those nostalgic for the old time parables of neoclassical writing, we must remind ourselves that scholars are not born to live an easy existence. We must respect, and appraise, the facts of life. (Samuelson 1966)

One might hope that such a definitive capitulation by as significant an economist as Paul Samuelson would have signalled a major change in the evolution of economics. Unfortunately, this was not to be. While many of the bishops have conceded that economics needs drastic revision, its priests preach on in a new millennium, largely unaware that they lost the holy war thirty years earlier.

The kernel

The term 'capital' has two quite different meanings in economics: a sum of money, and a collection of machinery. Economists assume that they can use the two terms interchangeably, and use the money value of machines as a proxy for the amount of machinery used in production. They prefer to abstract from the complexity that there are many different types of machines, many of which (such as, for example, blast furnaces) are solely suited to producing one particular commodity, and instead work with the generic term 'capital' – as if there is some generic productive substance which is just as suited to turning out sheep as it is to producing steel. For the economic theories of production and distribution to work, the behaviour of this hypothetical generic substance must be little different from the behaviour of the actual real world of many different machines.

However, a careful analysis of production as a system by which commodities are produced by combining other commodities and labour shows that the money value of machinery cannot be used as a proxy for the amount of machinery used in production. As a result, the economic theory of how commodities are produced is wrong, and the theory's argument that profit is a reward for capital's contribution to production is also wrong. This reinforces the observations made in the Chapter 5, that the distribution of income is not the result of impersonal market forces, but instead reflects the relative power of different social classes.

The roadmap

This quite difficult chapter begins with an outline of the economic theory of the production of commodities by 'factors of production', with its assumption that all machinery can be lumped into the aggregate called 'capital' and measured by the money value placed upon those machines. Then Sraffa's 'abstraction-free' analysis of production is outlined. It is shown that, rather than the rate of profit depending upon the amount of capital, as neoclassical economists argue, the measured amount of capital in fact depends upon the rate of profit.

Measuring capital

Though the war only began in earnest in 1960, the possibility of conflict was first flagged by Piero Sraffa in his 1926 paper 'The law of returns under competitive conditions' (discussed in Chapter 3). In passing, Sraffa observed that an essential aspect of the economic theory of production was the assumption that the interdependence of industries could be ignored. The problem was that this assumption was invalid when changes in one industry's output affected the costs of many other industries, which in turn determined the costs facing the first industry. As Sraffa put it,

> the assumption becomes illegitimate, when a variation in the quantity produced by the industry under consideration sets up a force

which acts directly, not merely upon its own costs, but also upon the costs of other industries; in such a case the conditions of the "particular equilibrium" which it was intended to isolate are upset, and it is no longer possible, without contradiction, to neglect collateral effects. (Sraffa 1926)

Sraffa spent the better part of the next thirty-five years turning this observation into a rigorous theoretical argument. The product was a book with the bland but descriptive title of The Production of Commodities by Means of Commodities, and the rather more revealing but still oracular subtitle of 'Prelude to a critique of economic theory'. Essentially, Sraffa provided the techniques needed to highlight fundamental internal inconsistencies in the economic theory of production.

This theory argues that commodities – everything from corn flakes to steel mills – are produced by 'factors of production'. These are normally reduced to just labour on the one hand, and capital on the other. This concept is normally embodied in a 'circular flow diagram' like that of Figure 1, which shows factors of production 'flowing' from households to the factory sector, and goods flowing from the factory sector to households.

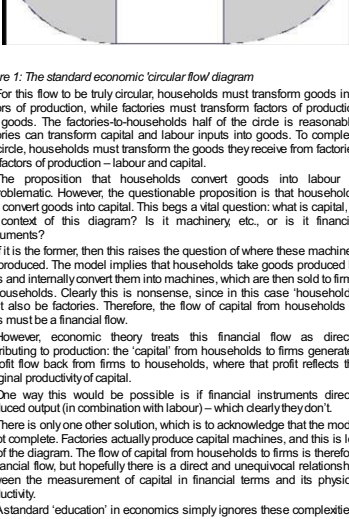


Figure 1: The standard economic 'circular flow' diagram

For this flow to be truly circular, households must transform goods into factors of production, while factories must transform factors of production into goods. The factories-to-households half of the circle is reasonable: factories can transform capital and labour inputs into goods. To complete the circle, households must transform the goods they receive from factories into factors of production – labour and capital.

The proposition that households convert goods into labour is unproblematic. However, the questionable proposition is that households also convert goods into capital. This begs a vital question: what is capital, in the context of this diagram? Is it machinery, etc., or is it financial instruments?

If it is the former, then this raises the question of where these machines are produced. The model implies that households take goods produced by firms and internally convert them into machines, which are then sold to firms by households. Clearly this is nonsense, since in this case 'households' must also be factories. Therefore, the flow of capital from households to firms must be a financial flow.

However, economic theory treats this financial flow as directly contributing to production: the 'capital' from households to firms generates a profit flow back from firms to households, where that profit reflects the marginal productivity of capital.

One way this would be possible is if financial instruments directly produced output (in combination with labour) – which clearly they don't.

There is only one other solution, which is to acknowledge that the model is not complete. Factories actually produce capital machines, and this is left out of the diagram. The flow of capital from households to firms is therefore a financial flow, but hopefully there is a direct and unequivocal relationship between the measurement of capital in financial terms and its physical productivity.

Astandard 'education' in economics simply ignores these complexities, and explains profit just as it explains wages: the payment to capital represents its marginal productivity. In particular, Figure 5.1 becomes Figure 2 here, with capital substituted for labour on the horizontal axis, and the rate of profit substituted for the real wage on the vertical.

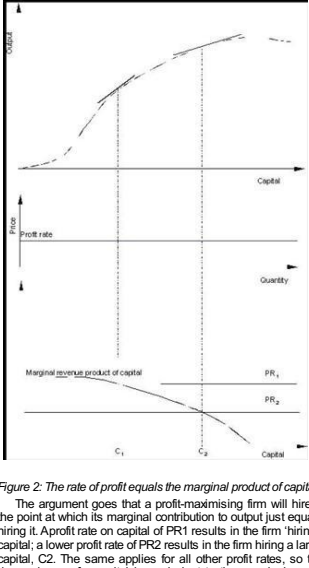


Figure 2: The rate of profit equals the marginal product of capital

The argument goes that a profit-maximising firm will hire capital up to the point at which its marginal contribution to output just equals the cost of hiring it. A profit rate on capital of PR1 results in the firm 'hiring' C1 units of capital; a lower profit rate of PR2 results in the firm hiring a larger amount of capital, C2. The same applies for all other profit rates, so that the firm's demand curve for capital is equivalent to the marginal product curve for capital.

The sum of all the individual demand for capital curves gives the market demand curve for capital. The point of intersection of this downward-sloping demand curve with the upward sloping supply curve yields the equilibrium rate of profit, as shown in Figure 3.

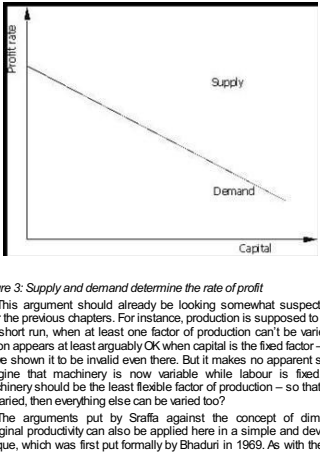


Figure 3: Supply and demand determine the rate of profit

This argument should already be looking somewhat suspect to you, after the previous chapters. For instance, production is supposed to occur in the short run, when at least one factor of production can't be varied. That notion appears at least arguably OK when capital is the fixed factor – though we've shown it to be invalid even there. But it makes no apparent sense to imagine that machinery is now variable while labour is fixed. Surely machinery should be the least flexible factor of production – so that if it can be varied, then everything else can be varied too?

The arguments put by Sraffa against the concept of diminishing marginal productivity can also be applied here in a simple and devastating critique, which was first put formally by Bhaduri in 1969. As with the labour market, the 'capital market' is a broadly defined 'industry': there would be thousands of products being lumped together into the general rubric of

'capital', and there is no industry which does not use some 'capital' as an input. This raises Sraffa's argument in Chapter 3, that a change in the price of such an input would affect numerous industries, and therefore alter the distribution of income. This is a similar point to that made earlier for the labour market, but it can now be put in a more explicit form.[46]

If we notionally divide all people into either workers or capitalists, then total income will be the sum of wages and profits. Profits in turn are the product of the rate of profit, times the amount of capital hired. Applying this at the level of the single firm, this gives us the relationship that:

Income equals
(a) the wage rate
multiplied by the number of employees plus
(b) the rate of profit
multiplied by the stock of capital

If we now consider changes in output (which we have to do to derive the marginal product of capital), then a rule of mathematics tells us that the changes in output have to equal the changes in wages and profits. Another rule of mathematics lets us decompose the change in profits into two bits: the rate of profit times the change in capital, and capital times the change in the rate of profit.[47] This yields the relationship that:

Change in income equals
(a) change in the wages bill (which we leave aggregated), plus
(b) change in profit (which we disaggregate)

Disaggregating changes in profit leads to the statement that:

Change in income equals
(a) change in the wages bill, plus
(b) the rate of profit
multiplied by the change in capital, plus
(c) the amount of capital
multiplied by the change in the rate of profit

At the level of the individual firm, economists assume that (a) and (c) are zero: a change in the firm's level of output caused solely by hiring more capital has no impact on either the real wage or the rate of profit. Thus the relationship can be reduced to:

Change in income equals
(a) change in wages [zero], plus
(b) the rate of profit
multiplied by the change in capital [one,[48] *plus*
(c) capital
multiplied by the change in the rate of profit [zero]

Cancelling out the terms we know are zero or one yields the desired relationship:

Change in output due to a change in capital (marginal product) equals the rate of profit

However, while this is a reasonable approximation at the level of the individual firm, it is not true at the level of the overall economy. There, any change in capital will definitely have implications for the wage rate, and for the rate of profit. Therefore the aggregate relationship is

Change in output due to a change in capital (marginal product) equals
(a) change in wages due to change in capital [non-zero], plus
(b) the rate of profit, plus
(c) the amount of capital
multiplied by the change in the rate of profit due to the change in capital [non-zero]

The rate of profit will therefore not equal the marginal product of capital unless (a) and (c) exactly cancel each other out.[49] Thus at the aggregate level, the desired relationship – the rate of profit equals the marginal product of capital – will not hold true. This proves Sraffa's assertion that, when a broadly defined industry is considered, changes in its conditions of supply and demand will affect the distribution of income.

A change in the capital input will change output, but it also changes the wage, and the rate of profit. These changes alter the distribution of income between workers and capitalists, and will therefore alter the pattern of demand. Exactly the same argument applies to wages, so that in general a person's income will not be equal to their marginal contribution to output. As

a result, the distribution of income is neither meritocratic nor determined by the market. The distribution of income is to some significant degree determined independently of marginal productivity and the impartial blades of supply and demand.

This adds what mathematicians call an additional 'degree of freedom' to the model of the economy. To be able to work out prices, it is first necessary to know the distribution of income; and there will be a different pattern of prices for every different division of the economic cake between workers and capitalists. There is therefore nothing sacrosanct about the prices that apply in the economy, and equally nothing sacrosanct about the distribution of income. It reflects the relative power of different groups in society – though it is also constrained by limits set by the productive system, as we will soon discuss.

This contradicts economic theory, which says that the distribution of income is uniquely determined by the market (via the mechanisms discussed in these two chapters), and therefore there's nothing that policy-makers can or should do to alter it. Instead, rather than prices determining the distribution of income as economists allege, the distribution of income determines prices. Within limits, the distribution of income is something which is determined, not by market mechanisms, but by relative political power.

Bhaduri's critique still accepts the assumption that it is possible to define a factor of production called capital. However, as I intimated above, the machinery aspect of the term 'capital' covers too great a multitude of things to be easily reduced to one homogeneous substance. It includes machines and the buildings that house them; trucks, ships and planes; oil wells, steel works and power stations. Each of these items itself consists of numerous other sub-assemblies which are themselves commodities. A truck contains an engine, which contains valves, springs and cables, the manufacture of which requires inputs from other types of capital, and so on.

The only thing that such disparate commodities obviously have in common is a price, and this is how economists would prefer to aggregate capital. But the price of a piece of capital should depend on the rate of profit, and the rate of profit will vary as prices change: there is an impossible circularity in this method of aggregation.

This problem was explicitly considered by Sraffa in his 1960 magnum opus. His purpose was to provide a firm foundation upon which a critique of the economic theory of production and income distribution could be built. He built his argument up stage by stage, with great care taken at each stage to make sure that the analysis was sound.

This meticulous method uncovered a number of paradoxes that invalidated the simplistic beliefs economists held about the relationship between productivity and income. Just as the peculiar conditions of 'production' of labour complicate the argument that the wage equals the marginal product of labour, so do the more conventional conditions of the production of capital disturb the argument that profit represents the marginal productivity of capital.

> *Note: the next section is possibly the most difficult part of this entire book. If you're satisfied with the debunking above, then you can skip this section for now and move to the next chapter. But I do recommend reading this section at some stage.*

The whole box and dice

Sraffa's technique was to eschew the initial aggregation of capital, and to say, in place of 'factors of production produce goods', that 'goods produce goods' – in concert with labour. Sraffa then used this 'assumption-free' model of production to show that the economic theories of price and of income distribution were invalid.

The essential point in his analysis was that capital does not exist as an easily definable entity, yet such an existence is necessary for the simple parable that profit represents the marginal productivity of capital to be true. He made this point by constructing a series of models that directly confronted the true complexity of a system of commodity production.

Sraffa built his models up very carefully, from a simple model with very little real-world realism to a more complex model which, with one exception, was a fairly realistic rendition of a market system of production.

The exception was that Sraffa considered an economy in equilibrium, when a real world economy is certain not to be in equilibrium. However, Sraffa's purpose was to critique economics on its own terms, and since economics assumes equilibrium, Sraffa made the same assumption. He took it to its logical conclusion, considering an economy which was not only in equilibrium now, but had been in equilibrium for the indefinite past.

Model one: production with no surplus

His first model was one in which the economy was just able to reproduce itself, and in which there was no 'fixed capital' – instead, all inputs were 'circulating capital' which are used up in each round of production.

In this economy, the output of each industry was just sufficient to supply

the demand for its output by itself and the other industries. Labour was not explicitly treated, but it was feasible to envisage that part of the inputs to an industry represented workers receiving a subsistence wage. Sraffa's example is shown in Table 1.

Industries	Wheat Input	Iron Input	Pig Input	Total Outputs
Wheat	240 qrs	12 ton	18 pigs	450 qrs
Iron	90 qrs	6 ton	12 pigs	21 ton
Pigs	120 qrs	3 ton	30 pigs	60 pigs
Total inputs	450 qrs	21 ton	60 pigs	

Table 1: Sraffa's hypothetical subsistence economy

In this hypothetical economy, combining 240 quarters of wheat, 12 tons of iron and 18 pigs in a production process results in an output of 450 quarters of wheat. Similarly, 90 quarters of wheat, 6 tons of iron and 12 pigs are used to produce 21 tons of iron; and 120 quarters of wheat, 3 tons of iron and 30 pigs are used to produce 60 pigs.

The total output of each sector just equals the amount of its output used to produce both its own output and that of all other sectors. Thus the total demand for wheat as an input is 450 quarters: 240 in wheat production, 90 in iron and 120 in pig production.

Sraffa posed the question of what would determine prices in this hypothetical economy, and the answer was not 'demand and supply', but 'the conditions of production': each sector's price had to enable it to just purchase its inputs. Specifying this for the wheat industry, this meant that 240 times the price of wheat, plus 12 times the price of iron, plus 18 times the price of pigs, had to just equal 450 times the price of wheat.

Similar equations applied for iron and pigs, and with three equations (the price equations for each sector) and three unknowns (the prices), there was one unique set of prices which made it possible for the economy to reproduce.[50]

Neoclassical economists might have endeavoured to find this set of prices by considering the demand curves for wheat, pigs and iron, and the supply curves for wheat, pigs and iron, and solving these to find the set of relative prices that equated supply and demand in each industry. However, in this context this would have been overkill: the only prices that work for this economy are those that enable each sector to buy its inputs.

Model two: production with a surplus

The next step towards realism was to consider an economy which produced a surplus: where at least one sector produced more of its output than was used up to produce itself and all other commodities. This step closer to a real market economy raises the issue of profits – which weren't an issue in the first model. For this economy to be in equilibrium, the rate of profit has to be the same across all sectors – even if only one sector produced a physical surplus. Otherwise, capitalists in sectors with a low rate of profit would be tempted to move to sectors with a high rate of profit, and the economy would not be in equilibrium. Sraffa used a 2-sector example, as shown in Table 2.

Industries	Wheat Input	Iron Input	Total Output
Wheat	280	12	575
Iron	120	8	20
Total Inputs	400	20	

Table 2: Production with a surplus

This economy uses 280 quarters of wheat and 12 tons of iron to produce 575 quarters of wheat; another 120 quarters of wheat and 8 tons of iron are used to produce 20 tons of iron. 175 bushels of wheat are produced over and above the 400 used in production, whereas the entire 20 tons of iron are used up in producing wheat and iron.

For a uniform rate of profit r to apply, the prices in this economy must be such that the 'money' value of inputs, multiplied by (1+r), must equal the money value of its outputs. For this example economy, the price ratio is 15 bushels of wheat for 1 ton of iron, and the uniform rate of profit is 25%.

Model three: production with a surplus and explicit labour

The economy above had to have labour in it, since nothing can be produced without labour.[51] However, this was not explicitly shown. The next model added further realism by showing that output was produced by combining both commodities and labour in a production process.

This introduces the wage as an additional unknown, and establishes the first element in Sraffa's critique of the economic theory of income distribution: rather than prices determining the distribution of income, the distribution of income between wages and profits must be known before

prices can be calculated.[52]

Sraffa then shows that there is an appropriate measuring stick (the 'standard commodity') which reveals a simple, linear relationship between the wage w, the actual rate of profit r, and the maximum feasible rate of profit for a given economy, R.[53] The wage w falls linearly as the rate of profit r rises towards its maximum value R.

The example economy in Table 2 has a maximum rate of profit of 25%, and results in the wage/profit function shown in Figure 4. If the wage w is .8 – which means that workers wages represent 80% of the surplus output of this economy – then the corresponding rate of profit r is 5%. This is shown numerically in Table 3.

Maximum R	25%
Wage (% of surplus)	Profit Rate
0%	25%
10%	23%
20%	20%
30%	18%
40%	15%
50%	13%
60%	10%
70%	8%
80%	5%
90%	3%
100%	0%

Table 3:

What this table says is that if workers, for example, get a zero wage, then all of the surplus goes to the capitalists, who then make a profit of 25%. If, however, workers get 10% of the surplus as their wage, then the rate of profit falls to 23% (rounded up). The same linear process continues right out to the point at which workers get 100% of the surplus, at which point capitalists get nothing and therefore have a rate of profit of zero. The same analysis is repeated graphically in Figure 4.

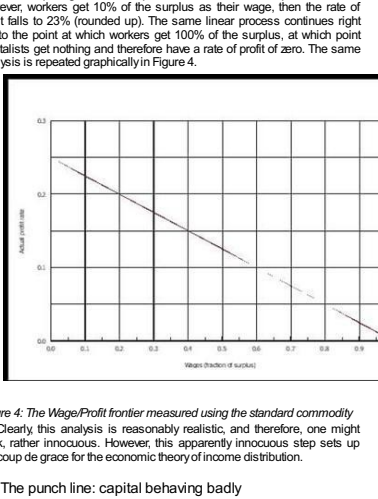


Figure 4: The Wage/Profit frontier measured using the standard commodity

Clearly, this analysis is reasonably realistic, and therefore, one might think, rather innocuous. However, this apparently innocuous step sets up the coup de grace for the economic theory of income distribution.

The punch line: capital behaving badly

The key concept in the neoclassical theory of income distribution is that factors get paid in accordance with their marginal contribution to output in the context of diminishing marginal returns. This means that as the supply of a factor increases, its return should fall.

The difficulty is, as alluded to earlier, that it is not easy to see how one can add units of capital together. Workers can be aggregated by adding up the number of hours they work – after notionally standardising for different levels of productivity by multiplying the hours of skilled labour by some amount to reflect higher productivity. Land can be aggregated by adding up acres – and again by adjusting numerically for varying degrees of fertility.

But machines have no apparent common property apart from price. This is in fact how economic theory aggregates capital, but this involves an obvious circularity, because the price of a machine reflects the profit expected from it, yet the rate of profit is the ratio of profit to price.

Sraffa proposed an ingenious and logically sound method of aggregation: to reduce capital to dated inputs of labour. The previous linear relationship between the wage and the rate of profit was an essential element in this analysis.

All items of capital are produced by other items of capital and labour. When an economy has been in equilibrium for the indefinite past, it is thus possible to regard the value of a machine as being equal to the value of the machines used to produce it, plus the value of the labour involved, times a rate of profit to reflect the passage of time. If we notionally treat the period of production as a year, then if the equilibrium rate of profit is 5%, 1.05 times the value of the inputs last year should equal the value of the machine this year.

The same argument applies to all the machines and labour inputs used to produce the inputs, and to all the machines and labour that produced them, and so on.

If we repeat this process, and each time reduce machinery inputs to the machinery and labour used to produce them, then we get a set of labour terms and a declining – but never zero – residual of machinery inputs. Each labour input is multiplied both by the wage, and by one plus the rate of profit raised to a power which reflects how many years ago the input was made.

If, for example, we are considering a machine manufactured 11 production periods ago, then this term will be the amount of direct labour bestowed in producing all the relevant components in the 12th year, times the wage, plus the capital input, all raised to the 12th power. It is therefore possible to substitute an expression in terms of labour for the capital inputs used up in producing a given commodity.[54]

We can now approximately express the value of a machine in terms of the sum of the value of the labour inputs used to produce it. Each element in this sum consists of a physical quantity of labour, multiplied by two terms: one representing the wage, and another representing the impact of accumulated profit over time.

The former term is a negative function of the rate of profit (as in Figure 4 and Table 4 above); the latter is as a positive function of the rate of profit, raised to a power. The former will fall in size as the rate of profit rises; the latter will rise, and it will also rise more for inputs made a long time ago.

This combination of opposing effects – one term that falls as r falls, the other that rises as r falls – evokes the possibility that one effect can prevail for a time, only to be overwhelmed by the opposite effect at a higher rate of profit. Therefore, the individual terms that interact to determine the value of an item of capital can rise for a while as the rate of profit rises, only to fall as the rate of profit rises still further.

This can be illustrated using Sraffa's example economy where the maximum rate of profit was 25%, and considering a machine which was made using 1 unit of labour as an input at some time in the past.

If the rate of profit was zero, then no matter how many years ago that machine was made, if a machine cost 1 (standard commodity) unit to make, its measured value would still be 1, as shown by the first row of Table 4. If the rate of profit was instead 1%, then the measured value of that machine falls to 0.96, if it is used today – reflecting the lower value of labour in terms of Sraffa's measuring stick.

Profit Rate	Years 0	1	2	3	4	5	10
0	1	1	1	1	1	1	1
0.01	0.96	0.97	0.98	0.99	1.00	1.01	1.06
0.02	0.92	0.94	0.96	0.98	1.00	1.02	1.12
0.03	0.88	0.91	0.93	0.96	0.99	1.02	1.18
0.04	0.84	0.87	0.91	0.94	0.98	1.02	1.24
0.05	0.80	0.84	0.88	0.93	0.97	1.02	1.30
0.1	0.60	0.66	0.73	0.80	0.88	0.97	1.56
0.2	0.20	0.24	0.29	0.35	0.41	0.50	1.24
0.21	0.16	0.19	0.23	0.28	0.34	0.41	1.08
0.22	0.12	0.15	0.18	0.22	0.27	0.32	0.88
0.23	0.08	0.10	0.12	0.15	0.18	0.23	0.63
0.24	0.04	0.05	0.06	0.08	0.09	0.12	0.34
0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4: The impact of the rate of profit on the measurement of capital

The value of the machine rises a bit if it was made 2 years ago, because its value is calculated to be 0.96 times 1 plus the rate of profit. This is 0.96 times 1.01, or roughly 0.97. This larger amount, though, is still less than 1, which would have been its value if the rate of profit had been zero. The same applies if the machine was used 2 periods ago, in which case its calculated value would be 0.98 – or 0.96, multiplied by 1.01 squared.

However, if the machine was produced 5 years ago, then its value in terms of the standard commodity rises to 1.01. This is because, while one part of the overall term has fallen to 0.96, the other has risen to 1.01 multiplied by itself 5 times – which roughly equals 1.05 – and 1.05 times 0.96 gives us 1.01.

The same effect applies across the row of the table, showing that as the rate of profit rises, the measured value of this capital input rises. The second term, 1.06, is 0.96 times 1.05 raised to the 10th; the third, 0.96 times 1.05 raised to the 15th; and so on.

The measured value of the machine therefore falls because of a higher rate of profit, but then rises if it was used many years ago. And the table has even more complications.

Notice that as we go down the table – so that the rate of profit increases – the value of a machine input today falls smoothly. However, the value of a machine applied 5 years ago rises for a while, but then falls. This accurate picture is a lot more complicated than economists assumed it to be, and these complications rule out the simple correspondence economists believed existed between the 'amount' of capital and the rate of profit.

The complications arise because the two different effects in Sraffa's accurate measure of capital don't cancel each other out. The first is the value of a wage unit, given the rate of profit r. On the first row, that is 1 (reflecting a zero rate of profit); on the second, 0.96 (at a 1% rate of profit); the third, 0.92 (at a 2% profit rate); and so on. But the second effect is 1+r, raised to a power of five, reflecting how many years ago the input was made. On the first row, that term is 1 – because the rate of profit is zero – and 1 times 1 is 1. On the second row, it is 0.96 times 1.05, which is 1.01 raised to the fifth power. This is roughly 1.01, so the measured value of the machine has risen. On the third row, it has risen further to 1.02 – which is 0.92 times 1.1, which is 1.02 raised to the fifth. On the fourth, it is roughly the same – 0.88 times 1.16, which is 1.03 raised to the fifth.

But by the time we get to a 10% rate of profit, the value goes down to 0.97: here we have 0.6 times 1.61, which is 1.10 raised to the 10th. The impact of the falling value of the first term now outweighs the impact of the rising value of the second. By the time we get to a rate of profit of 20%, the value of this machine (in terms of the standard commodity) has fallen to just 0.5, having been as high as 1.02 at lower rates of profit.

So the measured value of a machine rises and then falls as the rate of profit rises, and also rises and then falls as the time at which the machine was used to produce a commodity becomes further in the past.

This is not exactly how economists think about capital as a factor of production. They had hoped that the rate of profit would fall smoothly as the amount of capital used in production rose, so that capital, like labour, would manifest diminishing marginal productivity. But Sraffa instead showed that not only was there no uniform relationship between the rate of profit and the amount of capital, but also the direction of causation was the opposite of what economists wanted. Rather than the rate of profit depending on the 'amount' of capital, the measured amount of capital actually depended on the rate of profit. This makes it impossible to argue that the rate of profit is determined by the marginal productivity of capital, and so this second leg of the economic theory of income distribution collapses.

Not only that, but the perverse relationship that exists between the measurement of capital and the rate of profit is going to cause perverse effects in production. A rising rate of profit might for a while make one method of producing a commodity cheaper than alternatives, but then at a still higher rate of profit, it might make it more expensive.

Sraffa provides one illustration of this by comparing the price of two commodities which start out equal when the rate of profit is zero, and where one becomes more expensive than the other as the rate of profit rises, only to have the other become more expensive as the rate of profit rises further still. One product has relatively more 'direct labour' applied to its production in the recent past, while the other has more direct labour applied in the far distant past. Sraffa likens the latter to wine produced by being aged in a barrel; the former could be regarded as producing wine of identical quality using advanced chemical processes.[55] The latter process would be regarded as 'capital intensive', since so much machinery is used directly in its production, while the former would be called perhaps 'time intensive' (or labour intensive if you imagine the barrels being tended over the years by cellarmasters).

At a zero rate of profit, the cost of each wine equals simply the sum of the wages paid to produce the wine – and for both methods of production to exist in equilibrium, the cost of the two techniques must be identical.

As the rate of profit rises from zero to a moderate uniform rate, the far distant application of labour needed to produce the barrel has comparatively

little impact, so that the wine produced using modern technology is more expensive. In this range of the rate of profit, production using modern technology would cease, since it would be uncompetitive with wine produced using the aging process.

However, as the rate of profit becomes higher still, the effect of compounding the rate of profit on the making of the cask becomes enormous, so that the aged wine becomes more expensive than its mass-produced cousin. Mass production would take over again – we would switch back to the apparently more 'capital intensive' means of production.

Finally, when the rate of profit reaches its maximum value and wages fall to zero, the cost of wine falls to simply the cost of the irreducible commodity components (the original grapes, etc.), and the price of the two types of wine could again coincide.

Subsequent economists used Sraffa's building blocks to illustrate that a method of production could start out superior to all others at a zero profit rate, become less profitable than some other methods at a higher rate, only to once again become the most profitable at a higher rate still.

This phenomenon of 're-switching' destroyed the simple proposition that the rate of return on capital represented the marginal product of capital. If a particular production technique had lost primacy to others at one rate of profit, then it could not regain that primacy at a higher rate of profit still, unless for a period it benefited from increasing marginal product. But if marginal product could alternately rise and fall, then there was no necessity that the market for capital should be well behaved. Demand curves could slope up as well as down, supply curves down as well as up, and no unique equilibrium position could be defined.

The causes of this apparent paradox are that the concept of capital as a homogenous substance is an illusion, and that what is capital intensive depends on the rate of profit. If the rate of profit is low, then the labour embodied in an ancient wine barrel is of little consequence, and the process of aging wine may well appear to be labour-intensive. But if the rate of profit is high, then compounding of this high rate of profit makes that ancient wine barrel of great value – and the process could be described as capital-intensive. Rather than the rate of profit depending on the quantity of capital, the quantity of capital (in terms of its value measured in embodied labour value) depends upon the rate of profit.

The intricate and interdependent processes of production thus generate many opportunities for factor returns to move one way and then the other as factor intensities rise. There is therefore no consistent relationship between factor productivity and factor incomes. Instead, the distribution of income between wages and profits is largely independent of the system of production. The distribution of income is a social phenomenon.

Economists fought against this conclusion, but every apparent victory was shown to be invalid. Ironically, the rebuttals to economic rejoinders often showed that the only conditions under which the economic position could hold would be if the ratio of capital to output was the same in all industries. This is the same condition needed to make Marx's labour theory of value hold, yet the neoclassical revolution which gave us modern economic theory was supposedly free of the nonsense conditions needed by its Marxian rival.

So What?

Just as Chapter 5 showed that the wage can't be explained as the marginal product of labour, this chapter has established that economic theory cannot justify the existing rate of profit as somehow reflecting the marginal productivity of capital. Instead, the rate of profit reflects relative power in our society as well as the technical capabilities of factories and the success or otherwise of recent waves of investment. It is clearly possible for the rate of profit to be 'too high' or 'too low', but conventional economics is of no use in establishing either level.

Ignorance is bliss

Of course, the average economist would never tell you that economic theory had suffered such a devastating blow. This is because the average young economist doesn't even know that this intellectual bout took place – the concepts in this debate don't make it onto the curriculum for either undergraduate or postgraduate students. Older economists cannot avoid some knowledge of the war, but they either erroneously believe that their camp won, or they dismiss the issue completely.

Today, economic theory continues to use exactly the same concepts which Sraffa's critique showed to be completely invalid – capital as an amorphous mass that can be costlessly moved from producing any commodity to any other, whose return reflects its marginal productivity, and which can be aggregated by adding up its price times quantity.

There is no better sign of the intellectual bankruptcy of economics than this.

However, this madness is often justified by an appeal to a methodological precept that the absurdity of a theory's assumptions is irrelevant – all that matters is that the theory's predictions accord with reality. We now turn to consider this popular but false defence of economics.

7. There is madness in their method

Why assumptions do matter, and why economics is so different from the true sciences

Economics would have us believe that it is a science, fully able to stand tall beside the more conventional physical sciences and mathematics.

After the preceding chapters, you may be inclined to doubt that belief. Surely, whatever 'science' is, one might hope that it is undertaken with more impartiality, regard for the facts and logical consistency than economics has displayed.

However, the critiques of conventional economics which form the substance of this book were devised by critical economists (and sometimes, inadvertently, by conventional economists themselves) and some of these critiques have been acknowledged as valid by some conventional economists. There is also a small but robust minority working on other approaches to economic analysis, as you'll find in Chapter 14. There are thus some systematic and logical aspects to what economists in general do, which could qualify as scientific behaviour.

The position I favour is that economics is a science, but a rather pathological one. I am particularly critical of what has occurred since 1950, but I still hold out hope of better behaviour in the future.

But before better behaviour can take widespread root, economics will have to wean itself from a methodological myth. This is the proposition, first put by Milton Friedman, that a theory cannot be judged by its assumptions, but only by the accuracy of its predictions.

Leaving aside the question of whether economics has ever accurately predicted anything, the argument that "the more significant the theory, the more unrealistic [are] the assumptions" is simply bad philosophy.

The kernel

Have you heard the joke about the chemist, physicist and economist who get wrecked on a desert isle, with a huge supply of canned baked beans as their only food ? The chemist says that he can start a fire using the neighbouring palm trees, and calculate the temperature at which a can will explode. The physicist says that she can work out the trajectory of each of the baked beans, so that they can be collected and eaten. The economist says "Hang on guys, you're doing it the hard way. Let's assume we have a can opener." [56]

That assumption is not too different from the type of assumption that economists routinely make, and yet they defend themselves on the apparently convincing grounds that the assumptions don't matter – a theory can only be evaluated on the basis of the accuracy of its predictions.

This methodological defence is invalid, because it confuses 'negligibility' assumptions, which argue that some minor details can be ignored, with 'domain' assumptions, which determine the range of applicability of a given theory. Assumptions also do matter to economists, in that they genuinely believe that their theories describe reality, and they reject economic argument that is not based upon their preferred set of assumptions.

The roadmap

In this chapter I outline the paper in which Friedman introduced the notion that 'assumptions don't matter'. Following Musgrave, I classify assumptions under three headings: negligibility assumptions, domain assumptions, and heuristic assumptions. Friedman's paradoxical statement that "the more significant the theory, the more unrealistic the assumptions" is only partially true of the first class of assumptions, and manifestly untrue of the latter two classes. Finally, I detail the many ways in which assumptions do matter to economists.

A paradoxical proposition

There would be few if any academic economists who have not had a lecture disturbed by some recalcitrant student, interjecting that the assumptions of the model being discussed are unrealistic. Fortunately, there is a simple weapon at hand: an appeal to the authority of Milton Friedman that a theory can't be judged by its assumptions, but only by how well its predictions accord with reality.

In fact, Friedman's case went further: he argued that unrealistic assumptions were the hallmark of good theory. In what Paul Samuelson later dubbed 'the F-twist', Friedman argued that

> Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense). The reason is simple. A hypothesis is important if it "explains" much by little, that is, if it abstracts the common and crucial elements from the mass of

complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone. To be important, therefore, a hypothesis must be descriptively false in its assumptions; it takes account of, and accounts for, none of the many other attendant circumstances, since its very success shows them to be irrelevant for the phenomena to be explained.

> To put this point less paradoxically the relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic", for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions. (Friedman 1953)

The

The proposition that a theory is not regarded as a description of reality, but merely as a way of predicting the future, is known as 'instrumentalism'. This position is superficially appealing, and sufficiently persuasive to quieten the average interjector. It appears scientific, in that most scientists would admit that their theories can never exactly describe reality. It also implies a healthy dose of theoretical agnosticism, in that the economist is purportedly detached from her theory, and is only really interested in 'the facts'.

However, despite its superficial appeal, instrumentalism suffers from several flaws, which were clearly set out by the philosopher Alan Musgrave in 1981. Musgrave argued that there were three classes of assumptions, and that Friedman's dictum was only partially true in the least important of them.

Negligibility assumptions

Negligibility assumptions state that some aspect of reality has little or no effect on the phenomenon under investigation. Friedman's paper made heavy use of the example of a ball being dropped near the earth, which fell very nearly 'as if' it had been dropped in a vacuum. In this instance it was valid to assume that the ball was falling in a vacuum, since air resistance has negligible impact on the ball's flight. However, the same was obviously not true of a feather dropped under the same circumstances.

Friedman argued that though it was unrealistic to say 'assume the ball was dropped in a vacuum', the theory of gravity had great explanatory power: it explained much (the acceleration of bodies in free fall close to the earth) with very little (a gravitational constant and simple calculus). This theory should only be dropped in favour of another if a rival is at least as accurate and equally acceptable on other grounds, or "when there exists a theory that is known to yield better predictions but only at a greater cost" (Friedman 1953).

Musgrave argued that much of Friedman's musings were reasonable in this domain, but that even here his 'dialectical' proposition that "the more significant the theory, the more unrealistic the assumptions" is overblown. In fact, it is possible to rephrase these 'unrealistic' statements as 'realistic' ones: for example, it is realistic to say that air resistance is negligible for dense bodies falling from rest over short distances. As Musgrave put it, these assumptions:

> are not necessarily "descriptively false", for they do not assert that present factors are absent but rather that they are "irrelevant for the phenomena to be explained" . . . Galileo's assumption that air-resistance was negligible for the phenomena he investigated was a true statement about reality, and an important part of the explanation Galileo gave of those phenomena. (Musgrave 1981)

However, negligibility assumptions are the minnows of the assumptions family. Far more important are domain assumptions, and it is these to which rightly troubled students often object.

Domain assumptions

A domain assumption specifies the conditions under which a particular theory will apply. If those conditions do not apply, then neither does the theory.

An economic example of this is the assumption that risk can be used as a proxy for uncertainty – an assumption that permeates the conventional theories of macroeconomics and finance, which we will investigate in Chapters 9 and 10.

Risk applies to situations in which the regularity of past events is a reliable guide to the course of future events. Gambling gives us many such examples: if a tossed coin is seen to land showing heads roughly half the time, then you can reliably bet that there will be a 50:50 chance of heads in the future. If anyone bet you that heads would in future come up only 40% of the time, it would be sensible to take the bet. A risky event will have a probability associated with it, and a variance of outcomes around those probabilities, which can be reliably estimated using the techniques of statistics.

Uncertainty applies when the past provides no reliable guide to future

events. Though the fact that we cannot predict the future is the essence of the human condition, the very nebulousness of uncertainty means that many people – and certainly the vast majority of economists – have difficulty grasping the concept. As a result, they act as if the quantifiable concept of risk can be safely substituted for unquantifiable uncertainty.

A somewhat intimate example might illustrate the fallacy of identifying uncertainty with risk.[57] Imagine that you are very attracted to a particular individual, and that you know this person has gone out with 20% of those who have asked him or her out in the past. Does this mean that you have a 20% chance of being lucky if you 'pop the question'?

Of course not. Each instance of attraction between two people is a unique event, and the past behaviour of the object of your desires provides no guide as to how your advances will be received. How he or she will react cannot be reduced to some statistical prediction based on past apparent regularities. From your perspective, their reaction is truly uncertain – and this uncertainty is at the root of much of the angst that romantic attraction generates.

A similar observation can be made about each new business investment. Even if similar investments have been made in the past, the economic environment of a new investment differs from those which have gone before. Past trends therefore cannot be confidently extrapolated to predict future performance – but this procedure is the essential assumption behind using statistics to calculate risk.

The assumption that risk can be used as a proxy for uncertainty when evaluating investments is therefore unrealistic. A theory that makes such an assumption is quite clearly not better than an alternative one which does not – quite the opposite in fact. This assumption says that the domain of relevance of the theory is a world in which the future is simply subject to chance.

Since there is no such world, the domain of applicability of theories which make such an unrealistic assumption is 'nowhere'. Yet assumptions of this type abound in economic theory (especially, it must be said, in the work of Milton Friedman).

Such an assumption should be made only if it fits into Musgrave's third class, the heuristic assumption.

Heuristic assumptions

A heuristic assumption is one which is known to be false, but which is made as a first step towards a more general theory. Musgrave gives the example of Newton's assumption that the solar system consisted only of the sun and the earth. This gave rise to the theory that planets would follow elliptical orbits (which is a reasonable medium-term guide to actual planetary orbits in our solar system).

The next major step came with Poincar– in 1899, who tried to develop a formula describing planetary motion in a system with more than one planet. His proof that there was no such formula – and that the actual orbits would interact in wildly unpredictable ways – ushered in what is now known as 'chaos theory' or 'complexity theory' (though it lay dormant for 68 years until modern computers allowed its accidental re-discovery).

The modern theory of planetary behaviour now recognises that the stable orbits of our solar system can only have evolved – over an enormous period of time – from far less stable orbits, which must have led to collisions between proto-planets. It is now accepted that the moon, for example, was the product of a collision between another proto-planet and the early earth.

Collisions are not possible in a single-planet solar system – the kind of system that Newton assumed to derive his initial theory. Though that heuristic assumption was a major step in the development of the scientific mode of thinking about astronomy, dropping it led to a better theory, not a worse one.

When heuristic assumptions are made consciously by a theorist in the course of developing a theory, they are normally explicitly described as such. For instance, when developing the theory of relativity, Einstein at one point stated that the distance covered by a person walking from one side to the other of a moving train is equal to the sum of the distance covered by the train, and the width of the carriage. However, he continued that "We shall see later that this result cannot be maintained; in other words, the law that we have just written down does not hold in reality. For the time being, however, we shall assume its correctness" (Einstein 1916). When Einstein dropped this heuristic assumption, the theory of relativity was the result.

The greater realism at the heart of Einstein's theory transformed our understanding of reality, and dramatically expanded the physical and intellectual capabilities of our species. Yet if we accept Friedman's methodology, then we would have to argue that Einstein's theory was poorer than Newton's because it was more realistic.

In general then, and contrary to Friedman, abandoning a factually false heuristic assumption will normally lead to a better theory – not a worse one.

Judging the assumptions

Theories can therefore be evaluated by their assumptions to some

extent, if one has an intelligent taxonomy of assumptions. A theory may well draw power from 'unrealistic' assumptions if those assumptions assert, rightly, that some factors are unimportant in determining the phenomena under investigation. But it will be hobbled if those assumptions specify the domain of the theory, and real world phenomena are outside that domain.

These assumptions may be justified if they are merely heuristic devices used to simplify the process of deriving a more general theory – but only if that more general theory is in fact derived. Economists often imply, when they fob off some critical student, that the unrealistic assumptions in introductory economics courses are dropped in more advanced theory – which portrays these assumptions as heuristic tools. In fact, as preceding chapters have illustrated, the assumptions used in more advanced theory are often more unrealistic than those presented in introductory lectures.

Scientific realism vs instrumentalism

Musgrave also points out that most scientists reject an instrumental view of science in favour of 'scientific realism' – the belief that scientific theories should not merely predict reality but should, in some sense, represent it.

Ironically, this is actually the belief that most economists have about economic theory. Friedman's instrumentalism is little more than a smokescreen behind which to hide when one wishes to quell a budding class rebellion. It is often evident to the student objector that, though professing that the assumptions don't matter, his teachers continue to use the same small class of assumptions over and over again: rational utility-maximising individuals, profit-maximising firms, and a plethora of ancillary assumptions built on these foundations.

These assumptions are used because economists believe that these assumptions do capture essential elements of reality, and regard any theory which does not use these building blocks as 'unrealistic'. This belief is most clearly seen in the manner in which the 'Bibles' of economics, its academic journals, filter out papers that do not make this core set of assumptions.

Assumptions do matter – to economists

The proposition that assumptions don't matter implies that economists would be quite willing to accept a theory which assumed irrational behaviour if the model generated results which accorded with observation. It also implies that the development of economic theory would be driven primarily by the desire to produce theories that provide a closer fit to observed data.

Both these implications are strongly at variance with reality.

As any non-orthodox economist knows, it is almost impossible to have an article accepted into one of the mainstream academic economic journals unless it has the full panoply of economic assumptions: rational behaviour (according to the economic definition of rational!), markets that are always in equilibrium, risk as an acceptable proxy for uncertainty, and so on. When it comes to safeguarding the channels of academic advancement, little else matters apart from preserving the set of assumptions that defines economic orthodoxy.

Similarly, the development of economic theory over time has been propelled by the desire to make every aspect of it conform to the preferred economic model. Macroeconomics, when it first began, bore little resemblance to microeconomics. Fifty years later, macroeconomics is effectively a branch of microeconomics. As I outline in Chapter 9, a major factor behind this tribal coup was the belief that, regardless of its predictive validity, macroeconomics was unsound because its assumptions did not accord with those of microeconomics. It was therefore extensively revised, especially during the 1970s and 1980s, so that macroeconomic theory was more consistent with microeconomic assumptions. Far from assumptions not mattering to economists, assumptions in fact drove the development of economic theory.

Assumptions and logic

Assumptions matter in a more profound sense because, as this book shows, assumptions can be logically incoherent. For example, as discussed in Chapter 4 and the extension to it on the web, the economic model of the firm is internally contradictory. Economists wish to prove that perfect competition is superior to monopoly, but to prove this result, they must show that the marginal product of numerous small firms equals the marginal product of one large firm. The only situation in which this is possible is when marginal product curves are horizontal – but with horizontal marginal product curves, the output of a perfectly competitive industry is indeterminate (because a horizontal marginal cost curve can never intersect a horizontal marginal revenue curve). The assumptions of the theory are thus logically inconsistent.

A theory that contains logically inconsistent assumptions will be a bad theory – and, as this book shows, economics is replete with logical inconsistencies.

This is a science?

The behaviour of economists hardly fits the stereotype of scientists as dispassionate seekers of truth. But their behaviour does fit modern, sociological theories of how scientists behave.[58]

Briefly, these theories argue that each 'science' is as much a society as it is an intellectual discipline. A collection of scholars in a science will share a perspective on what defines their discipline, and what constitutes scientific behaviour. This shared mind-set includes core beliefs, which cannot be challenged without threatening your membership of the group (and hence your status as a scientist), ancillary beliefs which are somewhat malleable, a set of analytic techniques, and as yet unsolved problems to which these techniques should be applied. The core beliefs are known as the 'hard core' – since they cannot be altered without rejecting, in some crucial sense, the very foundations of the science. The ancillary beliefs are known as the 'protective belt', since their function is to protect the core beliefs from attack.

The scholars expect that their beliefs and techniques will be able to solve the outstanding problems, thus increasing the explanatory power of their science. If they fail, then the first response is to adjust the ancillary beliefs rather than the core propositions. Only when the problem proves both intractable and crucial is there any possibility that core beliefs will be abandoned, leading to the formation of a new school of thought – or the ascendancy of an existing rival school. While a school of thought is expanding the range of phenomena it can explain using its core beliefs – by experiments that confirm its predictions, or extensions of its theories to novel areas – then it is said to be a 'progressive' scientific research program which manifests a 'positive heuristic'. If, instead, experimental results contradict its predictions, and its theories are adjusted to rationalise these failures, then it is said to be 'degenerative' with a 'negative heuristic'.

It is possible for more than one such collection of scholars to exist in a science at any one time, so it makes sense to speak of schools of thought within a science. Each school of thought will compete with the others, emphasising their weaknesses and its own strengths.

Clearly this sociological description of a science fits the historical record of economics. At the beginning of the third millennium, there are at least five schools of thought. The neoclassical school is clearly dominant, but there are several other competing schools – in particular, the Post Keynesian, Austrian, and evolutionary schools of economics. Each is developing its own approach to explaining similar phenomena, and there is clearly a rivalry between the minority schools and neoclassical economics – the other schools criticise neoclassical economics while it largely ignores its rivals.

However, it might be thought that this provides a fairly demeaning perspective on science itself. Surely this behaviour is aberrant, and true sciences are beyond this petty bickering? No, strange as it may seem, a similar picture can be painted even of the queen of sciences, physics.

Quantum uncertainty?

In order to comprehend some of the bizarre results of experimental particle physics, most physicists argue that matter is in some sense 'probabilistic', and that the observer fundamentally affects reality. If an observer tries to 'tie down' one aspect of a particle – say, its location – then some other aspect becomes fundamentally unknowable. Physicists say that an elementary particle is always in a 'superposition' of both states, and testing for one leads to the other state resolving itself in a completely random way. The act of observing a particle thus directly – but unpredictably – alters its state. This is not because of any statistical properties of large numbers of electrons, but because randomness is an inherent feature of fundamental particles.

Two crucial aspects of this 'Copenhagen school' interpretation of quantum reality are (a) that particles can be treated as 'wave functions' in what is known as the wave-particle duality, so that a fundamental particle can be completely represented by its wave function; and (b) that there are two sets of physical laws, one which applies when there is no observer ('superposition') and one which exists when there is an observer.

The most famous popular representation of what this means, when put in terms of everyday objects, is 'Schrodinger's Cat'. This is a thought experiment in which a box contains a cat, a radioactive element, and a vial of poison. If the radioactive element emits a particle, the vial opens and the cat dies. If it doesn't, the cat lives.

What state is the cat in before an experimenter opens the lid to see whether it is alive or dead? In the Copenhagen school interpretation, the cat is in a superposition of being both alive and dead. The act of the observer opening the box resolves the cat into one or other state.

But this is not the only way to make sense of the experimental data. A rival interpretation, established by David Bohm, provides a completely deterministic interpretation, with none of the 'quantum uncertainty' of the Copenhagen school. It can explain the same experimental results as can the Copenhagen school – and some which it can't explain – without resorting to the apparently metaphysical position that the observer

somehow affects reality at the quantum level. In Bohm's theory, Schrodinger's cat is either alive and well if the radioactive element hasn't emitted a particle, or dead if it has, independent of the human observer who eventually opens the box to check.

How have physicists reacted to this co-existence of two rival explanations of reality? As the physicist David Albert sees it, in much the same way that economists have reacted to alternative schools of thought – by refusing to take them seriously. It is worth citing Albert at some length to show that, quite possibly, scientists in other disciplines are no different from economists when it comes to their reaction to intellectual challenges to accepted dogma:

> Despite all the rather spectacular advantages of Bohm's theory, an almost universal refusal even to consider it, and an almost universal allegiance to the standard formulation of quantum mechanics, has persisted in physics, astonishingly, throughout most of the past 40 years. Many researchers have perennially dismissed Bohm's theory on the grounds that it granted a privileged mathematical role to particles. The complaint was that this assignment would ruin the symmetry between position and momentumas if ruining that symmetry amounted to a more serious affront to scientific reason than the radical undermining, in the Copenhagen formulation, of the very idea of an objective reality. Others dismissed Bohm's theory because it made no empirical predictions (no obvious ones, that is) that differed from those of the standard interpretation – as if the fact that those two formulations had much in common on that score somehow transparently favoured one of them over the other. Still others cited "proofs" in the literaturethat no deterministic replacement for quantum mechanics of the kind that Bohm had already accomplished was even possible. (Albert 1994)

At this sociological level, therefore, economics appears to be very similar to the conventional sciences. However, even at this level, the current state of economics can be distinguished from, for example, the current state of astronomy.

A degenerate scientific research program

There was a time when the neoclassical school of economics was clearly progressive, while its main rival was clearly degenerate. When the neoclassical school coalesced in the 1870s in the works of Jevons, Menger and Walras, the preceding classical school was in crisis. The classical school always had a difficulty in explaining the relationship between what it called value, and prices; yet it insisted that value was in some way fundamental to the determination of price. This problem was accentuated by the work of the final member of the classical school, Karl Marx (the subject of Chapter 13).

At the same time, the neoclassical school was expanding its core belief that human behaviour was driven by the desire to maximise utility. This had developed from a guiding principle, in Bentham's hands, to a coherent theory of consumer and producer behaviour in the hands of Jevons, and to an explanation for the overall co-ordination of a market economy in Walras. At the turn of the 19th century, neoclassical economists were confident that their science could continue expanding its explanation of the economy. It was clearly then a progressive scientific research program.

Though the majority of economists still believe that this is the case today, there are manifest signs that this is no longer true. Instead, the theory today is degenerate: rather than expanding the range of phenomena it can explain, the leading edge of the theory is dominated by adjusting the protective belt of ancillary beliefs to defend the hard core beliefs from attack. For example, the Sonnenshein-Mantel-Debreu conditions (discussed in Chapter 2) are a way of maintaining the hard core belief that individual behaviour is driven by utility maximisation, despite the proof that individual preferences cannot be aggregated. Asimilar interpretation could be given of responses of neoclassical economics to the many logical problems documented in this book.

But the problems with economics go beyond just this, since if economics were as fully a science as astronomy, eventually its litany of failures would lead to at least a general acknowledgement of crisis.

The incredible inertness of economics

What makes economics different from and inferior to other sciences is the irrational tenacity with which it holds to its core beliefs in the face of either contrary factual evidence or theoretical critiques that establish fundamental inconsistencies in its intellectual apparatus.

The discovery, for example, that firms believe they experience constant or falling marginal costs (Eiteman & Guthrie 1952), and generally set prices by placing a markup on average cost, led not to the abandonment of the economic theory of price setting, but to a welter of papers arguing that in a competitive market, the effect of markup pricing was the same as if firms did consciously equate marginal cost to marginal revenue (Langlois 1989). On the same note, Sraffa's theoretical argument that diminishing marginal

returns were unlikely to occur in practice was ignored.

As a result, students at the end of the 20th century are receiving much the same instruction about how firms set prices as did their counterparts at the end of the 19th century.

Physical sciences hold onto their core beliefs with some tenacity, but nowhere near this much – even Albert's paper goes on to observe that "serious students of the foundations of quantum mechanics rarely defend the standard formulation anymore" (Albert 1994). As a result, revolutions in physical sciences – where one dominant paradigm is replaced by another – occur much more frequently than they do in economics. Often, these revolutions outpace the popular understanding of a science.

Astronomy provides an example of this. I expect that most lay people think that the dominant theory of how the universe came into being is the 'Big Bang'. In this theory, the universe originated in a 'quantum singularity' some 12–15 billion years ago. This explosion kick started matter and time, leading to the immense universe we observe today. Back in the 1950s, this theory won out against its rival, that the universe had always been in a 'steady state' of expansion.

The Big Bang was indeed the dominant theory for some time – until it was pointed out that, according to calculations from quantum mechanics, the Big Bang would have resulted in a universe consisting of a mere handful of elementary particles.

Arival theory then developed which argued that, for a substantial period of time, the laws of physics of the current universe did not apply. Matter, for example, could move much faster than the speed of light. This 'inflationary universe' theory has subsequently been embellished to predict that there are many universes – as opposed to the one universe postulated by the Big Bang.

The shifts from the Big Bang paradigm to the inflationary universe, to 'multiverses', are big ones conceptually. The first envisages a single finite universe, while the last muses that ours may be only one of many universes, each with different 'fundamental' physical laws. But the science of astronomy made this move over a period of about 20 years, and it continues to undergo development today. Now even the inflationary/multiverse theory is under challenge, as measurements imply that the rate of expansion of the universe is actually increasing with time.

Economics, in contrast, has had only one acknowledged revolutionary episode in the last century – the Keynesian revolution during the 1930s. Yet at the end of the 20th century, the dominant school of thought in economics retains almost nothing from that revolution, and in fact appears to be a direct descendant of pre-Keynesian neoclassical economics.

Think of the many revolutions in our understanding of the physical world which have occurred in the 20th century: from Newtonian to Einsteinian physics; from Mendelian genetics to DNA and the human genome; from determinism to chaos theory. Any scientist from the 19th century would be bewildered by what is commonplace today in his discipline – save an economist.

Why is economics so resistant to change? Is it because everything economists believed at the end of the 19th century was correct? Hardly, as this book shows. Instead, to understand the incredible inertness of economics, we have to consider an essential difference between social sciences in general and the physical sciences, and the thorny topic of ideology.

My kingdom for an experiment

In the 19th century, scientists and philosophers of science generally believed that what distinguished the social sciences from the physical sciences was that the latter could undertake experiments to test their theories, whereas the former could not. In the 20th century, Popper instead argued that the distinction between a science – like physics – and a non-science – like astrology – was not that one could undertake experiments and the other could not, but that one made falsifiable statements, while the other did not. Popper's distinction between science and non-science wasn't completely relevant to the 'experiments versus no experiments' distinction, but it did tend to play down the importance of experimentation in deciding what was and what was not a science.

The history of economics implies that Popper's distinction does not give sufficient attention to whether or not a falsifiable statement can in fact be experimentally falsified. For example, Milton Friedman is famous as the father of the now relatively defunct sub-branch of economics known as monetarism. One falsifiable statement he made was that inflation is caused by the government increasing the money supply more rapidly than the economy is going.

This implied that, to reduce inflation, all the government had to do was to increase the money supply more slowly than the economy was growing. This was the basis of the economic policies of Margaret Thatcher, yet eventually this approach was abandoned. One reason why was that the government was never able to meet its targets for the rate of growth of the money supply – it might aim to increase it by, say, 6%, only to see it grow by 11%. Also, the relationship between the three crucial variables in

Friedman's theory – the rate of inflation, the rate of growth of the economy, and the rate of growth of the money supply – was never as watertight in practice as it appeared to be in his theory.

You could thus argue that Friedman's statement – that inflation is caused by the government expanding the money supply faster than the rate of growth of the economy – had been falsified. Did this lead Milton and his supporters to abandon his theory? Of course not: monetarists instead argued that all sorts of attenuating features disturbed the results.

In other words, because the monetarist experiment in Great Britain wasn't a controlled experiment, monetarist economists could refuse to accept that their theory had been falsified.

The same observation can be made about Marxist economists, and their attitude towards the data on Marx's theory that the rate of profit would tend to fall, or the inevitability of socialism, and so on. In other words, this isn't just a disease of the political Right, but an endemic problem in economics: without the ability to undertake controlled experiments, statements which could be falsified will be unfalsifiable in practice. Economists of all persuasions are therefore liable to hang on to beliefs that they argue are scientific, but which in the end are ideological.

The experience of another social science, psychology, provides some support for the argument that the ability to undertake experiments is crucial to scientific progress. For much of the 20th century, psychology was dominated by the 'behaviourist' school. This school argued that an organism's behaviour had to be understood as a response to an external stimulus: it was 'unscientific' to postulate any unobservable mental processes of the organism which mediated between the stimulus and the response. To this school, complex behaviour – such as playing a piano – had to be understood as a chain of stimuli and responses. However, experiments showed that

> Even average pianists move their hands too quickly for the tactile information to pass along the sensory nerves to the central nervous system and for the command to move the hands to be sent down the motor nerves . . . Therefore, the behaviourist hypothesis that each new action is a response to an external stimulus is implausible. (Bond2000)

This and several other experimental falsifications of behaviourism led to its demise, and replacement by cognitive psychology, which accepts that "there are cognitive processes that determine our behaviourwhich we, as psychologists, must explain, even if they are not directly observable" (Bond 2000). Thus psychology, with the help of experiments, was able to undergo a revolution from one dominant school to another – while economics continues to be dominated by the same school (which, ironically, has a very behaviourist view of human behaviour). Unless it develops a means to undertake experiments to test rival theories, economics may be unable to break from the grip of ideology.

Equilibrium and an invisible ideology

Economics as a discipline arose at a time when English society was in the final stages of removing the controls of the feudal system from its mercantile/capitalist economy. In this climate, economic theory had a definite (and beneficial) political role: it provided a counter to the religious ideology that once supported the feudal order, and which still influenced how people thought about society. In the feudal system the pre-ordained hierarchy of king, lord, servant and serf was justified on the basis of the 'divine right of Kings'. The King was God's representative on earth, and the social structure which flowed down from him was a reflection of God's wishes.

This structure was nothing if not ordered, but this order imposed severe restrictions on the now dominant classes of merchants and industrialists. At virtually every step, merchants were met with government controls and tariffs. When they railed against these imposts, the reply came back that they were needed to ensure social order.

Economic theory – then rightly called political economy – provided the merchants with a crucial ideological rejoinder. A system of government was not needed to ensure order: instead, social order would arise naturally in a market system in which each individual followed his own self-interest. Smith's phrase 'the invisible hand' came along rather late in the process, but the notion played a key role in the political and social transformations of the late 18th and early 19th centuries.

An essential aspect of this market social order was equilibrium.

From the outset, economists presumed that the market system would achieve equilibrium. Indeed, the achievement of equilibrium was often touted as an advantage of the free market over any system where prices were set by fiat. Equilibrium was therefore an essential notion of the economic defence of capitalism: the equilibrium of the capitalist market would replace the legislative order of the now defunct feudal hierarchy.

More importantly, whereas the feudal order endowed only the well born with welfare, the equilibrium of the market would guarantee the best possible welfare for all members of society. The level of individual welfare would reflect the individual's contribution to society: people would enjoy the

lifestyle they deserved, rather than the lifestyle into which they had been born.

If, instead of equilibrium, economists had promised that capitalism would deliver chaos; if, instead of meritocracy, economists had said that the market could concentrate inequality, then economists could have hindered rather than helped the transition to capitalism (though they more likely would have been ignored).

By the middle of the 19th century, the transition to capitalism was complete: what was left of feudalism was a mere vestige. But rather than the promised equilibrium, 19th century capitalism was wracked by cycles and enormous disparities of wealth. A major depression occurred roughly every 20 years, workers' conditions would improve and then rapidly deteriorate, prices rise and then fall, banks expand and then collapse. New 'robber barons' replaced the barons of old. It appeared that, while promising a meritocratic equilibrium, capitalism had instead delivered unbalanced chaos. A new political challenge arose: that of socialism.

Once again, economics rose to the challenge, and once again equilibrium was a central tenet. This time the defence was mounted by what we today call neoclassical economics, since classical economics had been turned into a weapon against capitalism by the last great classical economist, Karl Marx.

In contrast to the hand-waving of Smith, the neoclassical economists of the late 19th century provided a substantive mathematical analysis of how equilibrium could be achieved by an idealised market economy, and how this equilibrium could be fair to all. However, unlike the earlier classical championing of capitalism, this technical edifice provided very little in the way of libertarian slogans for the battle against the ideology of socialism. Instead of arming capitalism's defenders with rhetoric to deploy against socialists, it gave birth to the academic discipline of economics.

Capitalism eventually transcended the challenge of socialism, with little real assistance from economic theory. But while the economics had little impact upon capitalism, the need to defend capitalism had a profound impact upon the nature of economic theory. The defensive imperative, and the role of equilibrium in that defence, cemented equilibrium's role as a core belief of economic theory.

At the beginning of the third millennium, there is no competing social system against which capitalism must prove its superiority. Feudalism is long dead, and those socialist societies which remain are either socialist in name only, or bit players on the world stage.

Today, most economists imperiously dismiss the notion that ideology plays any part in their thinking. The profession has in fact devised the term 'positive economics' to signify economic theory without any value judgments, while describing economics with value judgments as 'normative economics' – and the positive is exalted far above the normative.

Yet ideology innately lurks within 'positive economics' in the form of the core belief in equilibrium.[59] As previous chapters have shown, economic theory has contorted itself to ensure that it reaches the conclusion that a market economy will achieve equilibrium.[60] The defence of this core belief is what has made economics so resistant to change, since virtually every challenge to economic theory has called upon it to abandon the concept of equilibrium. It has refused to do so, and thus each challenge – Sraffa's critique, the calamity of the Great Depression, Keynes's challenge, the modern science of complexity – has been repulsed, ignored, or belittled.

This core belief explains why economists tend to be extreme conservatives on major policy debates, while simultaneously believing that they are non-ideological, and motivated by knowledge rather than bias.

If you believe that a free market system will naturally tend towards equilibrium – and also that equilibrium embodies the highest possible welfare for the highest number – then ipso facto, any system other than a complete free market will produce disequilibrium and reduce welfare. You will therefore oppose minimum wage legislation and social security payments – because they will lead to disequilibrium in the labour market. You will oppose price controls – because they will cause disequilibrium in product markets. You will argue for private provision of services – such as education, health, welfare, perhaps even police – because governments, untrammelled by the discipline of supply and demand, will either under or oversupply the market (and charge too much or too little for the service).

In fact, the only policies you will support are ones that make the real world conform more closely to your economic model. Thus you may support anti-monopoly laws – because your theory tells you that monopolies are bad. You may support anti-union laws, because your theory asserts that collective bargaining will distort labour market outcomes.

And you will do all this without being ideological.

Really?

Yes, really – in that most economists genuinely believe that their policy positions are informed by scientific knowledge, rather than by personal bias or religious-style dogma. Economists are truly sincere in their belief that their policy recommendations will make the world a better place for

everyone in it – so sincere, in fact, that they often act against their own self-interest.

For example, there is little doubt that an effective academic union could increase the wages paid to academic economists. If economists were truly self-motivated – if they behaved like the entirely self-interested rational economic man of their models – they would do well to support academic unions, since the negative impacts they predict unions to have would fall on other individuals (fee-paying students and unemployed academics). But instead, one often finds that economists are the least unionised of academics, and they frequently argue against actions that, according to their theories, could conceivably benefit the minority of academics at the expense of the greater community. However ideological economists may appear to their critics, in their hearts they are sincerely non-partisan – and, ironically, altruistic.

But non-partisan in self-belief does not mean non-partisan in reality. With equilibrium both encapsulating and obscuring so many ideological issues in economics, the slavish devotion to the concept forces economists into politically reactionary and intellectually contradictory positions.

Of course, if economists were right that equilibrium embodies the best possible outcome for the greatest number, then their apparently ideological policy positions would be justified – if the economy always headed back to equilibrium when disturbed from its Nirvana. In the next chapter, we'll put aside the critiques which establish that the building blocks of equilibrium are invalid, and instead ask whether economic equilibrium, as defined by economic theory, is in fact stable.

8. Let's do the Time Warp again

Why economics must finally treat time seriously

Forget everything you know about riding a bicycle, and imagine that someone who purports to be a 'bicycle guru' has convinced you that there are two steps to learning how to ride a bike. In step 1, you master balancing on a stationary bike. In step 2, you master riding a moving bike, applying the skills acquired at step 1.

After several difficult months at step 1, you would know that to remain upright, you must keep your centre of gravity directly above the point of contact between your wheels and the road.

Step 2 arrives. Applying the lessons in stage 1, you keep your bike at a perfect 90 degrees to the ground, balance against the uneven pressure of your legs, get up some speed and you're away.

So far, so good. But what if you want to change direction? The handlebars appear to provide the only means of turning, so you rotate them in the direction you wish to go – and fall flat on your face.

What went wrong with this apparently logical advice? 'Elementary, my dear Watson': the gyroscopic force which keeps you upright when a bike is moving simply doesn't exist when it is stationary. Manipulating this force is what enables you to turn a moving bike, and the lessons learnt in the static art of balancing a stationary bike are irrelevant to the dynamic art of actually riding one.[61]

Replace the bicycle with the economy, and the point still stands: the procedures which apply in a static economy are irrelevant to a dynamic, changing one; the forces which apply in a static economy simply don't exist in a dynamic one.[62] Lessons learnt from managing an economy in which processes of change either don't occur, or in which changes occur instantly, are irrelevant to an economy in which change does occur, and takes time to occur.

The kernel

Economic theory in general ignores processes which take time to occur, and instead assumes that everything occurs in equilibrium. For this to be allowable, the equilibrium of the dynamic processes of a market economy must be stable, yet it has been known for over 40 years now that those processes are unstable: that a small divergence from equilibrium will not set up forces which return the system to equilibrium. The dynamic path of the economy therefore cannot be ignored, and yet most economists remain almost criminally unaware of the issues involved in analysing dynamic, time-varying systems.

The roadmap

In this chapter I detail the roots of the economic propensity to ignore time, and to instead focus on what happens in equilibrium. Then I point out that economic research in the 1950s and 1960s established that the equilibrium of a market economy was unstable, so that the economy could never be in equilibrium. A brief discussion of chaos theory outlines the type of analysis which economists should undertake, but do not.

Cobwebs of the mind

Economic processes clearly take time, and yet economists don't consider time in analysing demand, supply, or any of their other key variables. For example, the quantity demanded of a commodity and the quantity supplied are both treated as functions of price, and the outcome is an equilibrium. quantity To illustrate what they believe will happen if the demand for a commodity rises, economists draw two demand curves and one supply curve, as shown in Figure 1, and then practise what they call comparative statics.

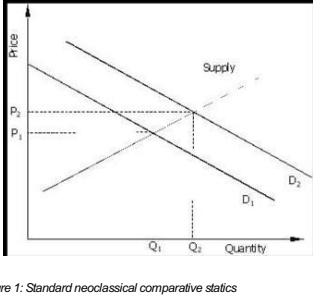


Figure 1: Standard neoclassical comparative statics

Before the increase in demand, the market is assumed to be in equilibrium with output at Q1 and price at P1. After the increase in demand (from D1 to D2), the market is assumed to be at the new equilibrium combination of Q2 and P2. Economists will then compare these two 'static' situations in terms of their impact upon consumer welfare, etc. – hence 'comparative statics'. The time path from one equilibrium to another is ignored.

What if the initial market price happens not to be the equilibrium price? Then demand and supply will be out of balance: if price exceeds the equilibrium, demand will be too low and supply too high. For equilibrium to be restored, this disequilibrium must set off dynamic processes in supply and demand which cause them both to converge on the equilibrium price. This dynamic process of adjustment will obviously take time. However, in general, economists simply assume that, after a disturbance, the market will settle down to equilibrium. They ignore the short-term disequilbrium jostling, in the belief that it is just a short-term sideshow to the long-run main game of achieving equilibrium.

A similar belief permeates even some of the alternative schools of economics. The dynamic process is ignored because it is believed to be a short-term, transitory phenomenon, and attention is focussed on the long-term, allegedly enduring phenomenon of equilibrium. As a result, time itself, the change in variables over time, and disequilibrium situations are all ignored. Even econometric programs which attempt to forecast the future value of macroeconomic variables such as output and employment assume that the current levels are equilibrium values, and they predict what the future equilibrium values will be.

Economics has invented numerous intellectual devices to enable itself to ignore time, and focus upon the equilibrium situations rather than consider the processes of change over time in an economy. One of these devices is one to which many budding students of economics initially object: 'all other things being equal', the 'ceteris paribus' assumption that nothing changes outside the single market being considered. This assumption lies behind the analysis of supply and demand, as shown in Figure 1 and discussed in chapters 2 to 4.

Such troubled students are reassured that at higher levels of analysis, this 'partial equilibrium' assumption is dropped for the more realistic proposition that all things are inter-related. However, rather than this more general analysis being more realistic, dynamic, and allowing for disequilibrium as well as equilibrium, it is in fact 'general equilibrium': a model of how all aspects of an economy can be in equilibrium simultaneously.

Budding economists who object to the assumption of ceteris paribus would walk away in disgust if they were immediately told of the assumptions needed to sustain the concept of general equilibrium. However, their fears assuaged by the promise of more realistic notions to come, they continue up the path of economic inculcation. By the time they confront general equilibrium in graduate education, they treat these assumptions and the analysis which goes with them as challenging intellectual puzzles, rather than as the asinine propositions they truly are. Normally, these students work at less rarefied levels of economic theory, and confidently presume that the leading lights of the profession will generalise the assumptions and solve the remaining puzzles.

As is so often the case with economic theory, the leading lights have done their job very well, but they have not delivered the goods expected of them by the troops. Instead, they have proven that, in general, general equilibrium is unattainable. Even economic models will not achieve general equilibrium, let alone the real economies that general equilibrium once purported to model. General equilibrium is at one and the same time the crowning achievement of economic theory and its greatest failure.

General equilibrium

In the late 19th century, three economists in different countries independently gave birth to the neoclassical school of thought: Jevons in England, Menger in Austria, and Walras in France. Today, Walras is the most exalted of these, because his model of general equilibrium set the mould by which economics has since been crafted.

Groping towards equilibrium

According to economic theory, equilibrium occurs in a particular market when demand at a given price equals supply at the same price. For equilibrium to occur in all markets simultaneously, the price in every market has to be such that demand and supply are equal in all markets. However, a change of price in one market will affect consumer demand in all other markets. This implies that a move towards equilibrium by one market could cause some or all others to move away from equilibrium. Clearly it is possible that this 'dance of many markets' might never settle down to equilibrium.

This will be especially so if trades actually occur at disequilibrium prices – as in practice they must, since who could ever know when one real world market was in equilibrium, let alone all of them simultaneously? A disequilibrium trade will mean that the people on the winning side of the bargain – sellers if the price is higher than equilibrium – will gain real income at the expense of the losers, compared to the alleged standard of equilibrium. This shift in income distribution will then affect all other markets, making the dance of many markets even more chaotic.

Walras provided a simple initial abstraction to sidestep this dilemma: he assumed that no trades take place until equilibrium is achieved in all markets. Having satisfied himself that, in the absence of trade, the jiggling of prices up and down would eventually converge to equilibrium, he extended the same faith to a system with production and exchange at disequilibrium prices.

Walras envisaged the market as being a huge, and very unusual, auction. The audience for this auction includes all the owners of the goods for sale, who are simultaneously the buyers for all the goods on sale. At a normal auction, the quantity of each commodity offered for sale is fixed. In Walras' auction, the total amount of each commodity is fixed, but sellers will offer anywhere from none to all of this for sale, depending on the price offered. The quantity offered rises as the price rises, and vice versa, with any amount not sold being taken back home by the seller for his/her own consumption (there are no stocks; everything is either sold or consumed by the producer).

The most peculiar features of Walras' auction market are that, rather than selling each commodity one at a time, the 'auctioneer' attempts to sell all goods at once; and rather than treating each commodity independently, this auctioneer refuses to accept any price for a commodity until supply equals demand for all commodities. In Walras' words:

> First, let us imagine a market in which only consumer goods and services are bought and sold . . . Once the prices or the ratios of exchange of all these goods and services have been cried at random in terms of one of them selected as numeraire, each party to the exchange will offer at these prices those goods or services of which he thinks he has relatively too much, and he will demand those articles of which he thinks he has relatively too little for his consumption during a certain period of time. The quantities of each thing effectively demanded and offered having been determined in this way, the prices of those things for which the demand exceeds the offer will rise, and the prices of those things of which the offer exceeds the demand will fall. New prices now having been cried, each party to the exchange will offer and demand new quantities. And again prices will rise or fall until the demand and the offer of each good and each service are equal. Then the prices will be current equilibrium prices and exchange will effectively take place. (Walras 1874)

This is clearly not the way markets work in the real world.[63] Nonetheless, this mythical construct became the way in which economics attempted to model the behaviour of real world markets.

Walras' auctioneer starts the market process by taking an initial stab at prices. These arbitrarily chosen prices are almost certainly not going to equate demand and supply for each and every commodity – instead, for some commodities, demand will exceed supply, while for others supply will exceed demand. The auctioneer then refuses to allow any sale to take place, and instead adjusts prices – increasing the price of those commodities where demand exceeded supply, and decreasing the price where demand was less than supply. This then results in a second set of prices, which are also highly unlikely to balance demand and supply for all commodities; so another round of price adjustments will take place, and another, and another.

Walras called this iterative process of trying to find a set of prices which equates supply to demand for all commodities 'tatonnement' – which

literally translates as 'groping'. He believed that this process would eventually converge to an equilibrium set of prices, where supply and demand are balanced in all markets (so long as trade at disequilibrium prices can be prevented).

This was not necessarily the case, since adjusting one price so that supply and demand are balanced for one commodity could well push demand and supply further apart for all other commodities. However, Walras thought that convergence would win out because the direct effects on demand – of increasing the price of a commodity where demand exceeds supply, which directly reduces demand – would outweigh the indirect effects of changes in demand for other commodities. In his words:

> This will appear probable if we remember that the change from p''b to p''b, which reduced the above inequality to an equality, exerted a direct influence that was invariably in the direction of equality at least so far as the demand for (B) was concerned; while the [consequent] changes from p'c to p''c, p'd to p''d,, which moved the foregoing inequality farther away from equality, exerted indirect influences, some in the direction of equality and some in the opposite direction, at least so far as the demand for (B) was concerned, so that up to a certain point they cancelled each other out. Hence, the new system of prices (p''b, p''c, p''d,) is closer to equilibrium than the old system of prices (p'b, p'c, p'd,); and it is only necessary to continue this process along the same lines for the system to move closer and closer to equilibrium. (Walras 1874 [1954]; emphasis added)

'Generalising' Walras

Walras' ruse, of an auctioneer who stopped any trades taking place until such time as demand equalled supply in all markets, was clearly artificial. However, it enabled economists to make use of the well-known and relatively simple techniques for solving simultaneous linear equations.

The alternative was to describe the dynamics of a multi-commodity economy, in which trades could occur at non-equilibrium prices in anywhere from a minimum of two to potentially all markets. At a technical level, modelling non-equilibrium phenomena would have involved nonlinear difference or differential equations. In the 19th century, the methodology for them was much less developed than it is now, and they are inherently more difficult to work with that simultaneous linear equations.[64]

Walras' auctioneer was therefore arguably a justifiable abstraction at a time when, as Jevons put it, it would have been "absurd to attempt the more difficult question when the more easy one is yet so imperfectly within our power" (Jevons 1871 [1911]).

But it suggests an obvious, dynamic, research agenda: why not see what happens whether the artifact of no non-equilibrium trades is dispensed with? Why not generalise Walras' general equilibrium by removing the reliance upon the concept of equilibrium itself? Why not generalise Walras by dropping the fiction that everything happens at equilibrium?

This potential path was, for economics, the path not chosen.

Instead, the neoclassical 'Holy Grail' became to formalise Walras' concept of equilibrium: to prove that general equilibrium existed, and that it was the optimum position for society.

Unfortunately, reality had to be seriously distorted to 'prove' that general equilibrium could be attained. But, for the reasons given in Chapter 7, economists would rather sacrifice generality than sacrifice the concept of equilibrium.

The pinnacle of this warping of reality came with the publication in 1959 of Gerard Debreu's Theory of Value, which the respected historian of economic thought Mark Blaug has described as "probably the most arid and pointless book in the entire literature of economics" (Blaug 1998). Yet this "arid and pointless" tome set the mould for economics for the next forty years – and won for its author the Nobel Prize for economics.

"The formal identity of uncertainty with certainty"

Walras' vision of the market, though highly abstract, had some concept of process to it. Buyers and sellers would haggle, under the guidance of the auctioneer, until an equilibrium set of prices was devised. Exchange would then take place, and those prices would also determine production plans for the next period. There is at least some primitive notion of time in this series of sequential equilibria.

No such claim can be made for Debreu's vision of general equilibrium. In this model, there is only one market – if indeed there is a market at all – at which all commodities are exchanged, for all times from now to eternity. Everyone in this 'market' makes all their sales and purchases for all of time in one instant. Initially everything from now till eternity is known with certainty, and when uncertainty is introduced, it is swiftly made formally equivalent to certainty. A few choice extracts give a clearer picture of Debreu's total divorce from reality.

> For any economic agent a complete action plan (made now for

the whole future), or more briefly an action, is a specification for each commodity of the quantity that he will make available or that will be made available to him, i.e., a complete listing of the quantities of his inputs and of his outputs

For a producer, say the jth one, a production plan (made now for the whole future) is a specification of the quantities of all his inputs and all his outputs The certainty assumption implies that he knows now what input-output combinations will be possible in the future (although he may not know the details of technical processes which will make them possible)

As in the case of a producer, the role of a consumer is to choose a complete consumption plan His role is to choose (and carry out) a consumption plan made now for the whole future, i.e., a specification of the quantities of all his inputs and all his outputs.

The analysis is extended in this chapter to the case where uncertain events determine the consumption sets, the production sets, and the resources of the economy Acontract for the transfer of a commodity now specifies, in addition to its physical properties, its location and its date, an event on the occurrence of which the transfer is conditional. This new definition of a commodity allows one to obtain a theory of uncertainty free from any probability concept and formally identical with the theory of certainty developed in the preceding chapters. (Debreu 1953; emphasis added)

I can provide no better judgment of the impact this brazenly irrelevant theory had on economics than that given by Blaug:

Unfortunately this paper soon became a model of what economists ought to aim for as modern scientists. In the process, few readers realized that Arrow and Debreu had in fact abandoned the vision that had originally motivated Walras. For Walras, general equilibrium theory was an abstract but nevertheless realistic description of the functioning of a capitalist economy. He was therefore more concerned to show that markets will clear automatically via price adjustments in response to positive or negative excess demand – a property that he labelled "tatonnement" – than to prove that a unique set of prices and quantities is capable of clearing all markets simultaneously. By the time we got to Arrow and Debreu, however, general equilibrium theory had ceased to make any descriptive claim about actual economic systems and had become a purely formal apparatus about a quasi economy. It had become a perfect example of what Ronald Coase has called "blackboard economics", a model that can be written down on blackboards using economic terms like "prices", "quantities", "factors of production", and so on, but that nevertheless is clearly and even scandalously unrepresentative of any recognizable economic system. (Blaug 1998)

A hobbled general

It is almost superfluous to describe the core assumptions of Debreu's model as unrealistic: a single point in time at which all production and exchange for all time is determined; a set of commodities – including those which will be produced in the distant future – which is known to all consumers; producers who know all the inputs that will ever be needed to produce their commodities; even a vision of 'uncertainty' in which the possible states of the future are already known, so that certainty and uncertainty are formally identical. Yet even with these breathtaking dismissals of essential elements of the real world, Debreu's model was rapidly shown to need additional restrictive assumptions – as discussed in Chapter 2. Rather than consumers being able to have any utility function consistent with what economists decreed as rational, additional restrictions had to be imposed which, as one economist observed, came "very close to simply assuming that the consumers in the aggregate have identical tastes and income" (Diewert 1977).

This was not the end of the restrictions. As Blaug observes above, Walras hoped to show that the process of tatonnement would lead, eventually to equilibrium being achieved, and that the same outcome would follow even if disequilibrium trading occurred. In mathematical terms, he hoped to show that general equilibrium was stable: that if the system diverged from equilibrium, it would return to it, and that if the process of tatonnement began with disequilibrium prices, it would eventually converge on the equilibrium prices. Debreu abandoned this aspect of Walras' endeavour, and focussed solely on proving the existence of general equilibrium, rather than its stability. But stability cannot be ignored, and mathematicians have shown that, under fairly general conditions, general equilibrium is unstable.

Positive prices and negative stability

Walras' assumption that the direct effects of the price change would outweigh the indirect effects – so that the process of tatonnement would converge on the set of equilibrium prices – was reasonable, given the state of mathematics at the time. However, mathematical theorems worked out in

the 20th century established that, in general, this assumption is wrong.

These theorems established that the conditions which ensure that an economy can experience stable growth simultaneously guarantee that Walras' tatonnement process is unstable (Blatt 1983). Therefore if the auctioneer's first stab at prices is only a tiny bit different from the set of prices which would put all markets in equilibrium, his next stab – derived by increasing prices for goods where demand exceeded supply, and vice versa – will be further away from the equilibrium set of prices. The process of tatonnement will never converge to the equilibrium set of prices, so if equilibrium is a prerequisite for trade, trade will never take place.

These theorems[65] are too complex to be conveyed accurately by either words or figures, but in keeping with the objectives of this book, I'll attempt an explanation. If you don't want to twist your mind around the mathematical concepts involved, then please skip to the following heading ('Atransitional methodology?').

The 'general equilibrium problem' is to find a set of prices which result in the amount consumers demand of each and every product equalling the amount supplied. Prices obviously have to be positive, as do the quantities demanded and the quantities produced.[66]

Before commodities can be demanded, they must be produced, and the means of production are simply other commodities. If the economy is going to last indefinitely, the system of production must be able to generate growth.

This can be described by a set of equations in which the prices are the variables, and the quantities required to produce each commodity are the coefficients. Asingle equation adds up the cost of inputs needed to produce a given commodity at a given price. There will be as many equations as there are commodities to produce.

It is then possible to separate the prices into a column of numbers called a vector, and the quantities into a square of numbers called a matrix – where, as noted earlier, every element is a positive number. The properties of this matrix can then be analysed mathematically, and its mathematical properties can be used to answer economic questions.

This matrix is known as a Leontief input-output matrix, after the Russian economist who first developed this method of analysis. The first row of such a matrix effectively says that 'a units of commodity a combined with b units of commodity b andz units of commodity z will produce 1 unit of commodity a'. It is the simplest method of describing a system of production, in that it implies that there is one and only one best way to make each commodity: no substitution of one technology for another is allowed.

While this is much simpler model of production than economists like to work with, it turns out that the properties of this very simple system determine whether the equilibrium of any more general model is stable. If this simple system can't guarantee stability, then no more complex system is going to either (this is a general property of dynamic models: the stability of the system very close to its equilibrium is determined by its 'linear' parts, and Leontief's matrix is the linear component of any more complex model of production).

There are two stability conditions on the simple Leontief system: the quantities produced each year have to enable the system to reproduce itself (this won't happen if, for example, the required inputs of iron for year 10 exceed the output of iron in year 9); and the prices must be feasible (the iron-producing sector can't depend on the price of some required input to producing iron being negative, for example).

It turns out that the first stability condition is governed by a characteristic of the input-output matrix, whereas the second stability condition is governed by the same characteristic of the inverse of that matrix. As with simple constants, a matrix and its inverse have, to some extent, opposite properties. Thus if you have a constant a which is less than 1, then a squared will be much less than 1, a cubed even more so, and higher powers of a will eventually converge to zero. However, the inverse of a, 1/a will be greater than 1, and powers of 1/a will blow out to infinity. If the stability of some system depends upon both a and the inverse of a being less than 1, then no number can fulfil both requirements, and the system is going to be unstable.

Since economic models are supposed to concern themselves with real economies, which can and do change in size, the general conclusion is that a real economy will never be in a state of general equilibrium. If economics is to have any relevance to the real world – if economics is even to be internally consistent – then it must be formulated in a way which does not assume equilibrium. Time, and dynamic analysis, must finally make an appearance in economic analysis.

A transitional methodology?

The founding fathers of economics had no problem accepting such a conclusion. In fact, to them, static analysis was merely a stop-gap measure, a transitional methodology which would be superseded by dynamic analysis as economics reached maturity. Jevons, for example, argued that "If we wished to have a complete solutionwe should have to treat it as a

problem of dynamics." But he instead pioneered static analysis because "it would surely be absurd to attempt the more difficult question when the more easy one is yet so imperfectly within our power". (Jevons 1871 [1911]).

Similarly, and at more length, Marshall noted that

> dynamics includes statics . . . But the statical solution . . . is simpler . . .; it may afford useful preparation and training for the more difficult dynamical solution; and it may be the first step towards a provisional and partial solution in problems so complex that a complete dynamical solution is beyond our attainment. (Marshall 1907)

At the end of the 19th century, J.B. Clark, the economist who developed the marginal productivity theory of income distribution (critiqued in Chapter 5), looked forward to the 20th century as the period during which economic dynamics would supplant economic statics:

> A point on which opinions differ is the capacity of the pure theory of Political Economy for progress. There seems to be growing impression that, as a mere statement of principles, this science will fairly soon be completeIt is with this view that I take issue. The great coming development of economic theory is to take place, I venture to assert, through the statement and solution of dynamic problems. (Clark 1898)

In this paper, Clark gave many good reasons why economics should be analysed using dynamics rather than statics. Foremost amongst these was that

> A static state is imaginary. All actual societies are dynamic; and those that we have principally to study are highly so. Heroically theoretical is the study that creates, in the imagination, a static society. (Clark 1898)

One century later, economic dynamics has indeed been developed – but not by the school to which J.B. Clark belonged. Instead, neoclassical economics – the dominant school of thought in economics which is the focus of this book – still by and large ignores the issue of time. Students are often told that dynamics is important, but they are taught nothing but statics. A typical macroeconomics textbook, for example, states that "the examination of the process of moving from one equilibrium to another is important and is known as dynamic analysis". However, it then states that "Throughout this book we will assume that the economic system is stable and most of the analysis will be conducted in the comparative static mode" (Taslim & Chowdhury 1995).

In their research, economists of many persuasions – not just neoclassical economists – use static analysis because they believe that equilibrium is the enduring state of the economy, while dynamics merely captures the transient moments between different equilibria. For example, a Sraffian economist defended static methodology in economics by arguing that

> 'static' analysis does not 'ignore' time. To the contrary, that analysis allows enough time for changes in prime costs, markups, etc., to have their full effects. (Steedman 1992)

As the previous sections have shown, this confidence that 'the end point of a dynamic process is the state of static equilibrium' is completely wrong. So too were the founding fathers of economics, in their belief that dynamic analysis, to quote Clark, "does not invalidate the conclusions of a static theory" (Clark 1898). But even if they were right, even if dynamic forces did lead, eventually, to static outcomes, it would still be invalid to model the economy using static techniques. Keynes put the case best in 1923, when he made his oft-quoted but rarely appreciated observation that "in the long run we are all dead". The full statement gives a rather better picture of his intent:

> But this long run is a misleading guide to current affairs. In the long run we are all dead. Economists set themselves too easy, too useless a task if in tempestuous seasons they can only tell us that when the storm is long past the ocean is flat again. (Keynes 1923)

Keynes was right: it is not valid to ignore the transient state of the economy. As Fisher later observed, equilibrium conditions in the absence of disturbances are irrelevant, because disturbances will always occur. Whether equilibrium is stable or not, disequilibrium will be the state in which we live.

We also live in a changing – and normally growing – economy. Surely we should be concerned, not with absolute levels of variables, but with their rates of change? Should not demand and supply analysis, for instance, be in terms of the rate of change of demand, and the rate of change of supply? Should not the outcome of supply and demand analysis be the rate of change of price and quantity over time, rather than static levels? Should not macroeconomics concern itself with the rate of change of output and employment, rather than their absolute levels?

Of course they should. As Keynes also once remarked, "equilibrium is blither". So why, fifty years after Keynes, are economists still blithering? Why do economists persist in modelling the economy with static tools when

dynamic ones exist; why do they treat as stationary an entity which is forever changing?

There are many reasons, but the main one, as outlined in the previous chapter, is the extent to which the core ideological beliefs of economists are bound up in the concept of equilibrium. As a by-product of this, economists are driven to maintain the concept of equilibrium in all manner of topics where dynamic, non-equilibrium analysis would not only be more relevant, but frankly would even be easier. This obsession with equilibrium has imposed enormous costs on economics.

Firstly, unreal assumptions are needed to maintain conditions under which there will be a unique, 'optimal' equilibrium. These assumptions are often justified by an appeal to Friedman's methodological '', but as Chapter 7 pointed out, this notion is easily debunked. However, most economists take it as an article of faith, with insidious results. If you believe you can use unreality to model reality, then eventually your grip on reality itself can become tenuous.

Secondly, as shown in this chapter, even the unreal assumptions of general equilibrium theory are insufficient to save it from irrelevance, since even the model of general equilibrium has been shown to be unstable, so that no modelled or real economy could ever be in a state of equilibrium. Many of those who pioneered general equilibrium analysis are grudgingly conceding that these results require economics to radically alter direction. But they are also quite aware that lesser economists are, as Alan Kirman put it, "not even concerned over the sea-worthiness of the vessel in which they are sailing" (Kirman 1989).

Thirdly, the emphasis on modelling everything as an equilibrium phenomenon has isolated economics from most if not all other sciences, where dynamics – and in particular evolutionary analysis – is now dominant. Economists are now virtually the only 'scientists' who attempt to model a real world system using static, equilibrium tools. As a result of this isolation, economists have been shielded from developments in mathematics and other sciences which have revolutionised how scientists perceive the world.

This isolation is to some extent fortuitous, because if economists really knew what is common knowledge in other sciences, then economists would finally have to abandon their obsession with equilibrium, and economics as outlined in this book would cease to exist. Most modern-day economists believe, as did the founding fathers of economics, that dynamic analysis would simply 'fill in the dots' between the static snapshots, thus replacing a series of still photographs with a moving picture. In fact, modern research in mathematics, physics, biology and many other disciplines has shown that dynamic analysis normally leads to results which contradict those of static analysis.

In the long run, we are all in the short run

Equilibrium can be the long run destination of the economy only if it is stable – if any divergence sets up forces which will return the economy to equilibrium. Even after the proofs of the instability of general equilibrium, most economists believe that this is a non sequitur: surely, the equilibrium of any real-world system must be stable, since if it were unstable, wouldn't it break down? John Hicks articulated this view when he criticised one of the earliest dynamic models developed by an economist. He commented that Harrod (1939)

> welcomes the instability of his system, because he believes it to be an explanation of the tendency to fluctuation which exists in the real world. I think, as I shall proceed to show, that something of this sort may well have much to do with the tendency to fluctuation. But mathematical instability does not in itself elucidate fluctuation. A mathematically unstable system does not fluctuate; it just breaks down. The unstable position is one in which it will not tend to remain. (Hicks 1949)

The modern discipline known colloquially as chaos theory has established that this belief, though widespread amongst economists today, is quite simply wrong. The equilibrium of a real-world system can be unstable without the system itself breaking down.

The first and best illustration of this occurred, not in economics, but in meteorology. I'll give a brief exposition of this model, because it illustrates several ways in which the conventional economic understanding of dynamics is profoundly wrong. But first, we need a brief technical interlude to explain the difference between the mathematical methods used in static analysis and those used in dynamics (you can skip to 'The weather and the butterfly' if you'd like to avoid mathspeak).

Straight lines and curved paths

What static analysis means in technical terms is that the equations most neoclassical economists (and many non-orthodox economists) use in their mathematical models are 'algebraic' rather than 'differential'.

Algebraic equations are simply larger and more complicated versions of the equations we all did at school in geometry when we were asked to work out the intersection of two lines. Given two equations for Y in terms of X, with different slopes and Y intercepts, we worked out the only X point where the

two formulas gave the same Y point. Continuing with the geometry analogy, most of the equations used by economists use only straight lines, rather than more complicated shapes like parabolas, etc. Algebraic techniques with these equations scale indefinitely – you can have equations with hundreds of 'straight lines' and still get unique solutions.

Differential equations, on the other hand, are more complicated descendants of the technique of differentiation, which you might have learnt if you did calculus at school or college. Rather than being expressed in terms of X and Y, these equations are expressed in terms of the rate of change of X and the rate of change of Y. While school calculus only dealt with 'the rate of change of Y with respect to X', differential equations typically are in terms of 'the rate of change of Y with respect to itself, other variables, and time'.

Most differential equation models also involve curved relationships between variables, rather than straight lines. A straight line is in fact the simplest type of relationship which can exist between two variables (other than that of no relationship at all). Straight line relationships in differential equation models with unstable equilibria lead to ultimately absurd outcomes, such as negative prices, or cycles which approach infinite amplitude as time goes on. Nonlinear relationships, however, result in bounded behaviour: the forces which repel the system when it is very close to equilibrium are eventually overwhelmed by attractive forces when the system is substantially distant from the equilibrium.

Unlike linear algebraic equations, nonlinear differential equations don't scale well. Only a very few simple nonlinear differential equations can be solved – the vast majority can't be solved at all. Once there are more than two variables in a system of nonlinear differential equations, there is in fact no analytic solution. Such systems must be simulated to see what is actually going on.

The weather and the butterfly

In 1963, the meteorologist E.N. Lorenz devised a simple mathematical model of turbulent flow in a weather cell, using a simplified version of a well-known mathematical model of turbulent flow. His model had just three equations, with three variables and three constants. The first (x) equation described the intensity of convective motion, the second (y) the temperature difference between ascending and descending columns of air, and the third (z) described the divergence from linearity of the temperature gradient across the weather cell.[67]

It would be hard to think of a simpler set of three equations, and yet the behaviour they generated was unbelievably complex. Figure 2 shows the time path of the east-west fluid displacement.

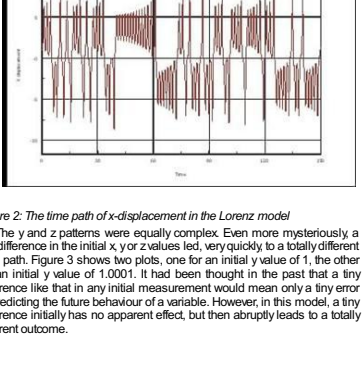


Figure 2: The time path of x-displacement in the Lorenz model

The y and z patterns were equally complex. Even more mysteriously, a tiny difference in the initial x, y or z values led, very quickly, to a totally different time path. Figure 3 shows two plots, one for an initial y value of 1, the other for an initial y value of 1.0001. It had been thought in the past that a tiny difference like that in any initial measurement would mean only a tiny error in predicting the future behaviour of a variable. However, in this model, a tiny difference initially has no apparent effect, but then abruptly leads to a totally different outcome.

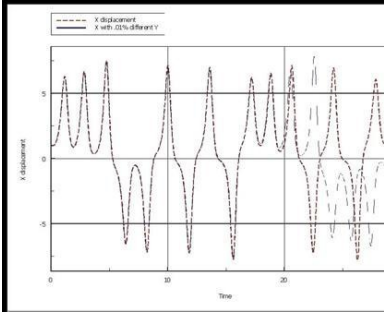


Figure 3: Sensitive dependence on initial conditions

Finally, though the pattern for any one variable appeared erratic, behind this apparent randomness lay a beautiful structure which is visible when the three variables are plotted on the one graph. Figure 4 shows the 'butterfly' behind the superficial chaos.

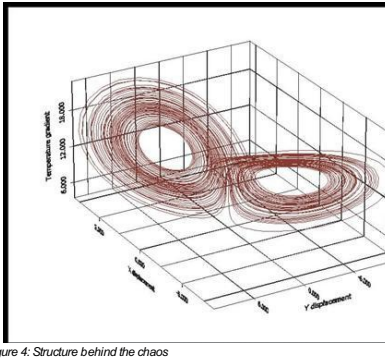
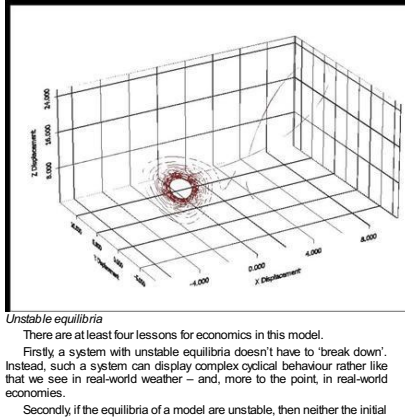


Figure 4: Structure behind the chaos

Detailed analysis of this system reveals that it has not one equilibrium, but three. More importantly all three equilibria are unstable. A slight divergence from any equilibrium causes the system moving to move away from it very rapidly. This is illustrated in Figure 5. A tiny divergence from one equilibrium point (of 0.00001%) leads to the system instantly being propelled from that equilibrium. It then approaches another, only to the flung off to a third. It orbits that equilibrium, only to be eventually repelled from it. Finally it approaches and then is repelled from the second equilibrium back towards the first.

Figure *5:*



Unstable equilibria

There are at least four lessons for economics in this model.

Firstly, a system with unstable equilibria doesn't have to 'break down'. Instead, such a system can display complex cyclical behaviour rather like that we see in real-world weather – and, more to the point, in real-world economies.

Secondly, if the equilibria of a model are unstable, then neither the initial nor the final position of the model will be equilibrium positions. The economic belief that dynamic analysis simply plots the movement between one equilibrium and another is therefore wrong. Instead, even simple dynamic models – of both weather and economics – will display 'far from equilibrium' behaviour. As a result, rather than equilibrium being where the action is, equilibrium tells you where the model will never be.

Thirdly, extrapolating from models to the real world, actual economic variables are likely to always be in disequilibrium – even in the absence of external shocks (or 'exogenous' shocks, as economists prefer to call them), which is the usual economic explanation for cycles – and the conditions which economists have 'proven' apply at equilibrium will therefore be irrelevant in actual economies. In this sense, equilibrium truly is, as Keynes put it, "blither". Static economic analysis therefore can't be used as a simplified proxy for dynamic analysis: the two types of analysis will lead to completely different interpretations of reality. In all such cases, the static approach will be completely wrong and the dynamic approach will be at least partially right.

Finally, even as simple a system as Lorenz's, with just three variables and three constants, can display incredibly complex dynamics because the interactions between variables are nonlinear (if you check the equations in footnote 7, you will see terms like 'x times y'). As noted earlier, nonlinear relationships in differential equation models can lead to complex but bounded behaviour.

From meteorology to economics

There are many models in economics which have properties akin to those of Lorenz's weather model – very few of which have been developed by orthodox economists. Most were instead developed by economists who belong to alternative schools, in particular complexity theorists and evolutionary economists. One of the best-known such models, Goodwin's model of cyclical growth, puts in mathematical form a model first suggested by Marx.

Marx argued that – in a highly simplified economy consisting of just capitalists and workers – there would be cycles in employment and income shares. A high rate of growth of output would lead eventually to high levels of employment. The high level of employment would encourage workers to demand large wage rises, which would eat into profits. The reduced level of profits would cause investment to decline, and growth would slow. The slower rate of growth would lead to increasing unemployment, which would in turn lead to workers accepting lower wages. Eventually the fall in workers' share of output would restore profit to levels at which investment would resume, leading to a higher rate of growth and higher employment levels. This would in time lead to wage demands once more, thus completing the cycle.

In mathematical form, this model reduces to two equations which are easily stated verbally:

$\dot{Y}$ The rate of change of workers' share of output equals workers' wage demands minus the rate of productivity growth;

$\dot{Y}$ The rate of change of employment equals the rate of growth of output, minus population growth and technological change.[68]

This mathematical model generates the cycle envisaged by Marx. Rather than converging to equilibrium values, workers' share of output and the rate of employment both cycle indefinitely, as shown in Figure 6.

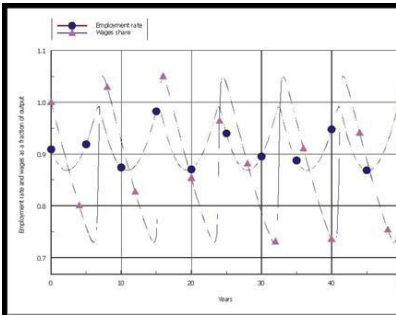


Figure 6: Cycles in employment and income shares

When wages share and employment are plotted against each other, the result is a closed loop, as shown in Figure 7. This is a far less complex structure than Lorenz's model, but it has one thing in common with it: the model does not converge to its equilibrium (which lies in the centre of the loop), but orbits around it indefinitely.

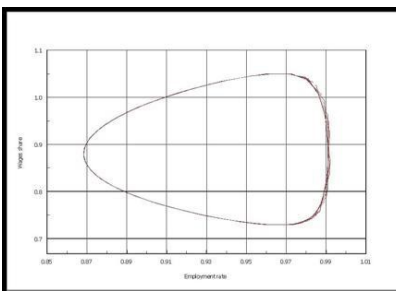


Figure 7: A closed loop in employment and wages share of output

It is also easily extended to capture more aspects of the real world, and when this is done, dynamic patterns as rich as those in Lorenz's model appear. For example, I have added equations to capture the accumulation of debt over a business cycle -- as businesses borrow money to finance investment during booms, and then have to repay the debt during slumps -- and the impact of government spending on the economy, which rises during slumps and falls during booms.[69]

Figure 8 shows the time path of employment, the debt to output ratio, and wages share of output in this more general model. As with Lorenz's model, the system is inherently cyclical: cycles in all three variables persist indefinitely, even without any 'exogenous shock' to disturb the equilibrium. As with Lorenz's model, the equilibrium of this system (there is only one) is unstable, yet rather than leading to breakdown, this instability generates sustained cyclical behaviour.

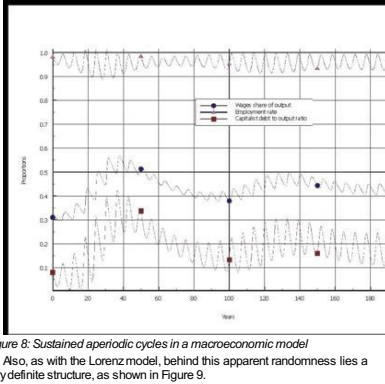
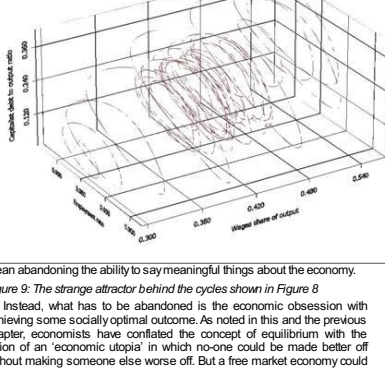


Figure 8: Sustained aperiodic cycles in a macroeconomic model

Also, as with the Lorenz model, behind this apparent randomness lies a very definite structure, as shown in Figure 9.

The point of this kind of modelling is not simply to draw pretty pictures, but to attempt to model real-world phenomena – from cyclones to the collapse of speculative manias. The key point is that these real-world phenomena simply cannot be modelled using 'comparative statics' or equilibrium – unless we are willing to believe that cyclones are caused by something 'exogenous' to the weather, and stock market bubbles are caused by something outside the economy. These models show that such phenomena can be modelled dynamically, so that abandoning static equilibrium analysis does not



mean abandoning the ability to say meaningful things about the economy.

Figure 9: The strange attractor behind the cycles shown in Figure 8

Instead, what has to be abandoned is the economic obsession with achieving some socially optimal outcome. As noted in this and the previous chapter, economists have conflated the concept of equilibrium with the vision of an 'economic utopia' in which no-one could be made better off without making someone else worse off. But a free market economy could never remain in an optimal position, because economic equilibria are unstable. The real question is whether we can control such an unstable system – whether we can constrain its instability within acceptable bounds.

This question was once at the heart of what is known as macroeconomics – the study of the entire economy and its control using government fiscal and monetary policy. Unfortunately, as we shall see in the next chapter, economists have formalised this once virile area of analysis into intellectual impotence.

9. The sum of the parts

Why Keynes's criticisms of conventional economics are still relevant today

Macroeconomics is the study of the economy at the aggregate level, with particular emphasis upon output, employment and inflation. The focus of this book, on the other hand, is overwhelmingly on what economists call 'microeconomics' – the study of disaggregated markets, where the theory attempts to derive market behaviour from the assumed characteristics of individual firms and consumers.

The major reason for this focus is because, by the end of the 20th century, macroeconomics had been almost totally subsumed by microeconomics.

The crowning glory of this process came with the development of 'representative agent' macroeconomics. This approach assumes that an economy can be modelled as if it consists of a single consumer/producer – the representative agent. As discussed in Chapter 2, economists invented this fiction because they had proven that it was impossible to treat collective welfare as simply the sum of the welfare of individuals (if individuals differed in tastes, or if tastes changed with income, which of course they do). The representative agent is clearly a fantasy, yet this fantasy became the standard way in which 'respectable' economists did macroeconomics.

It is stating the obvious to call the representative agent an 'ad hoc' assumption, made simply so that economists can pretend to have a sound basis for their analysis, when in reality they have no grounding whatsoever. Yet ironically, the prelude to this farce was the contention that macroeconomics was too ad hoc, because it was not grounded in microeconomic theory.

The kernel

Macroeconomics was once a vibrant and independent area of economic research from the study of individual markets, which is otherwise known as microeconomics. In spite of the manifest failings of microeconomic theory, economists have revised macroeconomics to make it more consistent with microeconomics. Macroeconomics today is just a branch of microeconomics, and is based on the proposition which Chapter 2 showed to be untenable, that society in the aggregate can be treated as 'one big consumer'. This process of decay was set in train firstly by Keynes's incomplete escape from conventional theory at the time he wrote The General Theory of Employment, Interest and Money, and accelerated by Hicks's dubious interpretation of Keynes as a marginalist. This interpretation ignored Keynes's argument that uncertainty was a fundamental aspect of economic reality with which economic theory had to grapple.

The roadmap

In this chapter I outline Keynes's critique of 'Say's Law', the argument that 'supply creates its own demand', and embellish it by comparing it to Marx's critique of the same proposition. I then argue that a key concept in neoclassical economics, 'Walras' Law', is simply Say's Law in a more formal guise. Keynes's and Marx's critiques of the conventional economics of their day are therefore still applicable to modern economics. Hicks's reinterpretation of Keynes as a 'marginalist' is debunked. Finally I detail Hicks's late realisation that his interpretation of Keynes was untenable once uncertainty was taken into account as a key determinant of the level of investment.

The self-equilibrating economy...

Macroeconomics only became a distinct sub-discipline within economics with the publication of Keynes's The General Theory of Employment, Interest and Money in 1936. This book was conceived and published during capitalism's greatest slump, the Great Depression, when America's output fell by 30% in 4 years, stock prices fell by 90%, commodity prices fell by almost 10% a year, and unemployment remained above 15% for a decade.

Prior to then, mainstream economists did not believe there were any intractable macroeconomic problems. Individual markets might be out of equilibrium at any one time – and this could include the market for labour or the market for money – but the overall economy, the sum of all those individual markets, was bound to be balanced.

The basis for this confidence was the widespread belief, amongst economists, in what Keynes termed Say's Law. As Keynes described it, this was the proposition that "supply creates its own demand" (Keynes 1936). Some economists dispute Keynes's rendition of Say's Law (Kates 1998), and I concur that in several ways Keynes obscured what Say actually meant. So it is appropriate to turn to the horse's mouth for a definition:

> Every producer asks for money in exchange for his products, only

for the purpose of employing that money again immediately in the purchase of another product; for we do not consume money, and it is not sought after in ordinary cases to conceal it: thus, when a producer desires to exchange his product for money, he may be considered as already asking for the merchandise which he proposes to buy with this money. It is thus that the producers, though they have all of them the air of demanding money for their goods, do in reality demand merchandise for their merchandise. (Say 1821)

Say's core proposition is that overall balance is assured because "[t]he sale of goods and services to the market is the source of the income from which purchases are financed" (Kates 1998). This, according to the 'classical' economists from whom hoped to distinguish himself [70] meant that there could never be a slump due to an overall deficiency in demand. Instead, slumps, when they occurred, were due to sectoral imbalances.

If the demand for one market – such as labour – was too low relative to supply, this was because demand exceeded supply in one or more other markets. The solution was for sellers in the market suffering from excess supply – workers – to accept a lower price for their commodity.

Money was also treated as a commodity in the pre-Keynesian model, and it was possible that, at some point in time, many people would want to hold money and very few would want goods. There could then be a serious slump, as producers of goods found that people did not want to part with their money. Physical commodity markets and the labour market could then be in excess supply – with unsold goods and unemployed workers – but this would be because of the excess demand for money, and not because of any overall deficiency of aggregate demand. In the aggregate, demand and supply would be in balance.

Keynes's attempt to refute this notion was, to put it kindly, rather confusing. It is therefore not altogether amazing that economists believed he either misunderstood this concept, or that he intended to refute the clearly incorrect belief that overall balance meant that there could never be involuntary unemployment – whereas Say's Law allowed for involuntary unemployment as a by-product of sectoral imbalances.

The upshot is that the essence of Say's Law lives on in modern economics, though it now goes under the more respectable name of 'Walras' Law' (or in some circles, 'Say's Principle'). Its modern definition is that 'the sum of all notional excess demands is zero', and this proposition is accepted as valid – indeed as irrefutable – by modern-day economists.

However, I argue that this is precisely the concept which Keynes intended to refute.

Say no more?

The modern argument [71] starts from the proposition that, on the average, agents in a market economy are neither thieves (who want to take more than they give) nor philanthropists (who want to give more than they get). Therefore the normal agent will intend to have balanced supplies and demands: the value of what she wishes to sell will equal the value of what she wishes to buy, so that 'the sum of her notional excess demands is zero'.

The excess demand for any single product by a single agent can be positive – so that the agent wishes to be a net buyer of that product – or negative – so that the agent wishes to be a net seller. However, in sum, her excess demands will be zero.

This balance at the level of the individual agent necessarily carries over to the aggregate of all agents: if the intended excess demands of each individual agent sum to zero, then the intended excess demands of all agents sum to zero.

However, this identity of aggregate supply and aggregate demand at the overall market level doesn't necessarily translate to identity at the level of each individual market. In particular, as noted earlier, it is possible for excess demand for money to be positive – in which case commodity markets would be 'glutted'. Excess demand for labour can also be negative – the supply of workers can exceed the demand for them – so that there will be involuntarily unemployed workers (and also a notional excess demand for the products that the unemployed workers intended to buy).

These two circumstances are both explanations of a depression. The former would involve a 'rising price' for money – or in other words 'deflation' as the money price of all other commodities fell. The latter would involve a falling price for labour – falling wages. However, both these forces would make a depression a temporary phenomenon. As Dixon puts it:

> [F]ollowers of Walras would say that involuntary unemployment cannot persist in a market economy with flexible wages and prices. They would argue that if the commodities market has excess demand then the prices of commodities will tend to rise and this will tend to reduce the level of excess demand in that market. In the labour market, where there is excess supply, they would assert that money wages will tend to fall. The joint effect of the rising price together with a falling money wage is that the real wage will tend to drop thus

reducing (and eventually removing entirely) the excess supply in the labour market.

As a consequence of the above, many would see the pronouncements of Keynes that the economy could find itself with an excess supply of labour and yet, in all (other) respects be in "equilibrium", as being in conflict with Walras' Law and therefore wrong or "bad" in theory and so inadmissible. (Dixon 2000)

Keynes's critique

Keynes's divided all output into two classes: consumption and investment. If the economy was in equilibrium, then Say's Law would argue that excess demand for consumption goods would be zero, and likewise for investment goods.

Keynes then imagined what would happen if demand for consumption goods fell, so that excess demand for consumption goods was negative (supply exceeded demand).[72] Say 's Law would argue that demand for investment goods would rise to compensate: notional excess demand for investment goods would be positive.

However, as Keynes argued extensively throughout the General Theory, demand for investment goods is driven by expectations of profit, and these in turn depend heavily upon expected sales to consumers. A fall in consumer demand now could lead entrepreneurs to expect lower sales in the future – since in an uncertain environment "the facts of the existing situation enter, in a sense disproportionately, into the formation of our long-term expectations; our usual practice being to take the existing situation and to project it into the future" (Keynes 1936). Dampened expectations would therefore lead entrepreneurs to reduce their demand for investment goods in response to a reduced demand for consumer goods. Thus a situation of negative excess demand for consumer goods could lead to a state of negative excess demand for investment goods too – a general slump.

This clearly contradicts Walras' Law. Since economists regard Walras' Law as irrefutable, this led some economists to ridicule Keynes's argument, and others to attempt to find how Keynes's argument could be reconciled with Walras' Law. The most widely accepted reconciliation was achieved by Robert Clower and Axel Leijonhufvud.

Say's Principle

Clower and Leijonhufvud asserted that Keynes and Walras were compatible because Walras' Law applied effectively only in equilibrium. Out of equilibrium, though the sum of notional excess demands was still zero, the sum of effective demands could be negative.

For example, if there was negative excess demand in the labour market – so that some workers were involuntarily unemployed – then it didn't help that these unemployed workers wanted to buy commodities. Without employment, their notional demands remained just that. Though they might want to buy commodities, without a wage their notional demand had no impact upon actual sales of commodities. Actual negative excess demand for labour might therefore not be balanced by actual positive excess demand for commodities, so that overall, the sum of excess demand could be negative. Keynes was vindicated as a disequilibrium theorist.[73] and Keynes and Walras were reconciled.

But were they? Later in his life, Keynes indicated that he rejected the very basis of Walras' Law– the proposition that the sum of notional excess demands is zero – when he praised the author of what he had once described as an "obsolete economic textbook which I know to be not only scientifically erroneous but without interest for the modern world" (Keynes 1925): Karl Marx.

The circuit of capital

Marx's critique of Say's Law, which so excited Keynes in his later years, went to the heart of Walras' Law (and Say's Law). Marx rejected Say's initial proposition that "[e]very producer asks for money in exchange for his products, only for the purpose of employing that money again immediately in the purchase of another product" (Say 1821). Instead, Marx pointed out that this notion asserted that no-one in a market economy wished to accumulate wealth. If there was never any difference between the value of commodities someone desired to sell and buy on the market, then no-one would ever desire to accumulate wealth. But an essential feature of capitalism is the existence of a group of agents with precisely that intention.

Believers in Walras' Law might find these agents rather bizarre, since in their terms these agents are 'thieves', who wish to take more than they give. However, far from being bizarre, these agents are an essential part of a market economy. They are known as capitalists. Far from their behaviour being peculiar in a market economy, it is in fact the essence of capitalism. Whereas both Say's Law and Walras' Law asserts that people simply desire to consume commodities, Marx asserted that an essential aspect of capitalism is the desire to accumulate. He derided Say's belief that the

ultimate objective of every agent in a market economy was simply consumption – which is still generally accepted by economists today, as well as the economists of Marx's time – as an ideologically convenient but misleading fiction which obscures the actual dynamics of capitalism:

> It must never be forgotten, that in capitalist production what matters is not the immediate use-value but the exchange-value, and, in particular, the expansion of surplus-value. This is the driving motive of capitalist production, and it is a pretty conception that – in order to reason away the contradictions of capitalist production – abstracts from its very basis and depicts it as a production aiming at the direct satisfaction of the consumption of the producers. (Marx 1861)

Capitalists are clearly fundamental to capitalism, and their behaviour directly contradicts the Walras' and Say's Law presumption that every agent's intended excess demand is zero. As Marx put it:

> The capitalist throws less value in the form of money into the circulation than he draws out of it . . . Since he functions . . . as an industrial capitalist, his supply of commodity-value is always greater than his demand for it. If his supply and demand in this respect covered each other it would mean that his capital had not produced any surplus-value . . . His aim is not to equalise his supply and demand, but to make the inequality between them . . . as great as possible. (Marx 1885)

The dilemma for Marx was to explain how this inequality could be achieved without 'robbing' other participants in the market, and without violating the principle that commodities were bought and sold at fair values. His solution points out the fallacy underlying the economist's superficially appealing arguments in Say's and Walras's Law.

This was that the market process had to include a production stage where the value of output exceeded the value of inputs – in Marx's terms, a surplus is produced. The capitalist pays a fair price for his raw materials and a fair wage to his employees. They are then combined in a production process which generates commodities for sale. The commodities are then sold for more than the cost of the raw materials and workers' wages, yielding a profit. The profit allows the capitalist to fulfil his desire to accumulate wealth, without robbing any other market participants, and without having to buy commodities below their value and sell them above it.

Say's Law and Walras' Law, on the other hand, envisage an exchange-only economy: an economy in which goods exist at the outset, but where no production takes place. The market simply enables the exchange of pre-existing goods. In such an economy, surplus in Marx's sense would be impossible. Equally, if one agent desired to and did accumulate wealth, that would necessarily involve theft in the Say's Principle sense. However, this condition does not hold when we move from the fiction of an exchange-only economy to the reality of a production and exchange economy. With production, it is possible for agents to desire to accumulate wealth without therefore aspiring to be thieves.

Marx formalised this analysis in terms of two 'circuits', the 'Circuit of Commodities' and the 'Circuit of Capital'.

In the Circuit of Commodities, people come to market with commodities, which they exchange for money in order to buy other commodities. Marx stylised this as C – M– C:

Commodity→Money→Commodity

Though Marx discussed various ways in which this circuit could fail – due primarily to delays between the sale of one commodity and the purchase of the next – generally speaking it obeys Walras's Law. Each 'agent' desires to convert commodities of a given value into different commodities of equivalent value.

However, in the Circuit of Capital, people came to market with money, with the intention of turning this money into more money. These agents buy commodities – specifically, labour and raw materials – with money, put these to work in a factory to produce other commodities, and then sell these commodities for (hopefully) more money, thus making a profit. Marx stylised this as M– C – M+:

Money→Commodity→More money

The complete circuit, and the one which emphasises the fallacy behind Walras's Law, was M– C(L,MP) . . . P . . . C+c – M+m:

Money→Labour and means of production . . . Production . . . Different commodities, of greater value than paid for the labour and means of production→ Sale of commodities to generate more money

This circuit specifically violates Say's Principle and Walras' Law. Rather than simply wanting to exchange one set of commodities for another of equivalent value, the agents in this circuit wish to complete it with more wealth than they started with. If we focus upon the commodity stages of this circuit, then as Marx says, these agents wish to supply more than they demand, and to accumulate the difference as profit which adds to their wealth. Their supply is the commodities they produce for sale. Their

demand is the inputs to production they purchase – the labour and raw materials. In Say's Principle's terms, the sum of these, their excess demand, is negative. When the two circuits are added together, the sum of all excess demands in a capitalist economy is likewise negative.

Taking this logic one step further, Minsky pointed out that the combination of a market economy in which, in an accounting sense, there is a buyer for every seller, and yet in which growth occurs over time, necessitates the existence of credit and debt:

> If income is to grow, the financial markets, where the various plans to save and invest are reconciled, must generate an aggregate demand that, aside from brief intervals, is ever rising. For real aggregate demand to be increasing, . . . it is necessary that current spending plans, summed over all sectors, be greater than current received income and that some market technique exist by which aggregate spending in excess of aggregate anticipated income can be financed. It follows that over a period during which economic growth takes place, at least some sectors finance a part of their spending by emitting debt or selling assets. (Minsky 1963 [1982])

Marx's perspective thus integrates production, exchange and credit as holistic aspects of a capitalist economy, and therefore as essential elements of any theory of capitalism. Conventional economics, in contrast, can only analyse an exchange economy in which money is simply a means to make barter easier.

Say's Principle, which insists that the sum of all notional excess demands is zero, is a model of a capitalist economy without production and, most importantly, without capitalists.

Walrasian rejoinders?

There are a number of objections which economists could make to Marx's model.

Firstly, Marx's circuits clearly cover not one market, but two: one when the capitalist buys his inputs, the other when he sells his outputs. Since these are two distinct markets in time, there is no reason, even under Walras' Law, why demands in one should equal supplies in the other. However, in each market, Walras' Law will apply.

Secondly, it is incorrect to conflate the exchange process with the production process. It is quite possible that agents could purchase inputs to production in one market, then combine them in production, produce a larger value of commodities and subsequently bring those commodities to sale at a subsequent market.

Thirdly, Marx's notion of a surplus implies that there are some commodities which can be purchased and, through production, turned into a larger value of commodities. This implies a 'free lunch'. If such a possibility ever existed, it would have long ago been 'arbitraged' away by the price of these commodities rising, or the price of the outputs falling.

Fourthly, Marx neglects the concept of a rate of time discount. Though some agents may appear to want to accumulate over time, if we discount future incomes to reflect the fact that the commodities that income will enable you to buy will be consumed in the future, then overall these agents are simply maintaining their level of satisfaction over time.

Taking the first and second hypothetical objections together, one of the strengths of Marx's approach is that his model covers a process through time, rather than merely considering an instant in time. In reality, at the aggregate level, exchange and production occur simultaneously. Factories are continuously producing commodities, sales rooms continually moving recently produced stock, workers are being paid wages, and spending them on consumer goods. Marx's circuits analysis captures the organic nature of the production and exchange processes of a market economy, whereas the economic approach artificially separates these into distinct stages.

This organic approach therefore enables Marx to consider the economy as a dynamic process, in which growth is an integral aspect of a capitalist economy. As part of this process, there are some agents who are continually accumulating wealth (when economic conditions are favourable), and others who are continually simply maintaining their level of economic well-being (though they can also gain in wealth if real wages are increasing, and if the wage exceeds subsistence).

Walras' Law, on the other hand, is best suited to the economic irrelevance of an exchange-only economy, or a production economy in which growth does not occur. If production and growth do occur, then they take place outside the market, when ironically the market is the main intellectual focus of neoclassical economics. Conventional economics is thus a theory which suits a static economy – and which can only be adapted to a dynamic economy with great difficulty, if at all – when what is needed are theories to analyse dynamic economies. Marx's 'through time' model of circuits is thus better suited to the analysis of a market economy than the 'moment in time' model of Walras' Law.

Marx's model of capitalist expectations is also far more valid than Walras'. A capitalist might well have his purchases and supplies balanced in any one market, as Walras' Law requires. However, purchases in this

period are undertaken with the intention of selling a greater quantity in the next market. Marx's notion of the circuit of capital thus provides a link between one 'market period' and the next, which Walras' Law does not.

Since Say's Law and Walras' Law are in fact founded upon the hypothesised state of mind of each market participant at one instant in time, and since at any instant in time we can presume that a capitalist will desire to accumulate, then the very starting point of Say's/Walras' Law is invalid. In a capitalist economy, the sum of the intended excess demands at any one point in time will be negative, not zero. Marx's circuit thus more accurately states the intention of capitalists by its focus on the growth in wealth over time, than does Walras' Law's dynamically irrelevant and factually incorrect instantaneous static snapshot.

The arbitrage argument highlights the difference between neoclassical theory and Marx's theory of value.[74] which I discuss in more detail in Chapter 13 (where, I had better point out, I reject the 'labour theory of value' which conventional Marxists regard as integral to Marx's analysis). Neoclassical theory basically argues that the average rate of profit is driven down to the marginal productivity of capital, so that profit simply reflects the contribution which capital makes to output. This rate of profit is then called 'normal profit', treated as a cost of production, and notionally set as the zero mark. Only profit above this level, called super-normal profit, is formally acknowledged in the theory, and in the pervasive theory of perfectly competitive equilibrium, super-normal profit is zero, so that profit fails to appear as a variable in economic theory.

The notion that profit is determined by the marginal product of capital was debunked in Chapter 6. Marx's theory of value, on the other hand, sees profit as a surplus of sales over the cost of production, allows for a positive rate of profit, and makes the rate of profit an integral part of the theory of production and exchange.

The time discount argument, that people are simply maintaining their level of satisfaction over time, has problems on at least two fronts. Firstly, it is very hard to believe, for example, that Bill Gates feels that his level of wealth in 2000 was equivalent to his wealth in 1970. Successful capitalists would clearly feel that they have gained in wealth over time – and unsuccessful capitalists would definitely know that they have lost. Secondly, all this argument does is move the zero position when calculating whether someone is accumulating, staying the same, or losing out. If the normal rate of time discount is, say, 2%, then anyone who is accumulating wealth at more than 2% per annum is increasing their wealth – the sum of their time-discounted excess demands is negative.

So what?

The Walrasian argument that the sum of all excess demands is zero provides an apparent centre of gravity for the economy. Rather like a see-saw, if one sector of the economy is down, then another sector is necessarily up. Furthermore, economists postulate that there are countervailing forces at work: a 'down' sector will have the price of its output driven down, thus increasing demand and restoring balance, and vice versa for an 'up' sector. The see-saw will ultimately return to balance.

A negative sum for aggregate excess demand removes that centre of gravity. Instead of the economy behaving like a see-saw where the pivot is carefully placed at the centre of gravity, it behaves like one where the pivot is instead off-centre, and can move abruptly one way or the other. A down sector is not necessarily offset by an up sector, so that, contrary to Walras' Law, one sector can remain down for an indefinite period.

In particular, a general slump is feasible. As Keynes argued, a decline in spending on consumption by consumers could lead investors to also reduce their demand for investment goods, so that the economy could remain in a situation of inadequate excess demand.

The key destabilising force is investment. As both Keynes and Marx emphasise, investment is undertaken not for its own sake, but to yield a profit. If expectations of profit evaporate, then so too will investment spending, and the economy can be thrown into a general slump. Equally, if expectations of profit become too euphoric, investment can be overdone, and the economy can be thrown into an unsustainable boom – in that the profits expected by the investors will not be realised, and the boom will give way to bust. The non-Say's/Walras' Law vision of the economy shared by Marx and Keynes thus accords with the manifest instability of the macroeconomy, whereas Walras' Law asserts that, despite appearances to the contrary, the macroeconomy really is stable.

Unfortunately, this perspective on Keynes's General Theory was buried beneath economists' mistaken belief that Walras' Law was incontrovertible. By forging a reconciliation between Keynes and Walras, the resulting 'Keynesian economics' was not Keynes's economics at all. It is little wonder that this Keynesian 'straw man' was so easily deconstructed by its conservative critics.

Say no more!

Marx provided a far clearer and more incisive critique of Say's Law than did Keynes, and when Keynes finally became aware of it, he endorsed

Marx's argument wholeheartedly Keynes also provided an eloquent explanation of why this shallow, simplistic notion held, and continues to hold, such a strong grip upon the minds of economists:

> That it reached conclusions quite different from what the ordinary uninstructed person would expect, added, I suppose, to its intellectual prestige. That its teaching, translated into practice, was austere and often unpalatable, lent it virtue. That it was adapted to carry a vast and consistent logical superstructure, gave it beauty. That it could explain much social injustice and apparent cruelty as an inevitable incident in the scheme of progress, and the attempt to change such things as likely on the whole to do more harm than good, commended it to authority. That it afforded a measure of justification to the free activities of the individual capitalist, attracted to it the support of the dominant social force behind authority. (Keynes 1936)

However, Keynes continued, this contrariness had, by the time of the Great Depression, led to a diminution of economics in the eyes of the public:

> But although the doctrine itself has remained unquestioned by orthodox economists up to a late date, its signal failure for purposes of scientific prediction has greatly impaired, in the course of time, the prestige of its practitioners. For professional economists, after Malthus, were apparently unmoved by the lack of correspondence between the results of their theory and the facts of observation – a discrepancy which the ordinary man has not failed to observe, with the result of his growing unwillingness to accord to economists that measure of respect which he gives to other groups of scientists whose theoretical results are confirmed by observation when they are applied to the facts. (Keynes 1936)

Despite Marx's and Keynes's critiques, Say's Law and Walras' Law lived on, and still dominate economic thinking today. The attempts by well-meaning Keynesian economists to reconcile Keynes with Walras' Law thus robbed Keynes of a vital component of his argument, making 'Keynesian economics' a severely emasculated version of Keynes's thought. But this was far from the only way in which Keynesian economics became a travesty of Keynes's original vision.

Hamlet without the prince

Rather like the Bible is for many Christians, the General Theory is the essential economics reference which few economists have ever read.

There are many reasons for this. One is that the General Theory is a difficult book.

There are at least two roots to this difficulty. The good reason is that Keynes was so much more insightful than most other economists that the concepts in the General Theory are difficult for more ordinary mortals to grasp. The bad reason is that, as Keynes himself acknowledged, the book was replete with concepts from the very school of economics which he was hoping to overthrow. As cited previously in the Preface, Keynes observed that

> The composition of this book has been for the author a long struggle of escape, and so must the reading of it be for most readers if the author's assault upon them is to be successful – a struggle of escape from habitual modes of thought and expression. The ideas which are here expressed so laboriously are extremely simple and should be obvious. The difficulty lies, not in the new ideas, but in escaping from the old ones, which ramify, for those brought up as most of us have been, into every corner of our minds. (Keynes 1936)

Another bad reason for not reading it is laziness: it is much easier to read the 'Reader's Digest' version given in a textbook than it is to slog through the unabridged original. As a result, many economists were inclined to rely upon summaries, rather than reading the original. Keynes obliged by providing his own summary, of just 15 pages, in 1937.

Keynes and uncertainty

The key concept in Keynes's summary was the impact of uncertainty upon investment, and therefore upon economic analysis.

Investment was undertaken to augment wealth, and yet the outcome of any investment depended upon economic circumstances in the relatively distant future. Since the future could not be known, investment was necessarily undertaken on the basis of expectations about an uncertain future.

Keynes was at pains to distinguish the concept of uncertainty from the simpler concept of risk.

Risk occurs when some future event can only be one of a number of already known alternatives, and when there is a known history of previous outcomes which enables us to assign a definite probability to each possible outcome. A dice roll is an example of risk. The dice can only land on one of six sides, and therefore only one of six numbers will turn up. If it is a fair dice, each number has a 1 in 6 chance of turning up. The theory of

probability can then be used to help predict the chances of various patterns of numbers occurring in future rolls of a dice.

Uncertainty is fundamentally different, and it has proven to be a difficult concept for economists before and after Keynes to grasp. Keynes gave several examples. Neither roulette, nor life expectancy, nor even the weather qualified. Instead, uncertainty referred to such things as the chance that war might break out (this was in 1937, not long before Chamberlain's 'Peace in our time' deal with Hitler), the rate of interest twenty years in the future, or when some invention would become obsolete. I gave a more personal and I hope evocative example of uncertainty in Chapter 7.

Probability theory cannot be used to help guide us in these circumstances because there is no prior history to go on, and because the outcomes are not constrained to any known finite set of possibilities. As Keynes put it, "About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know" (Keynes 1937).

Faced with this uncertainty, we develop conventions to help us cope. We assume that "the present is a much more serviceable guide to the future than a candid examination of past experience would show it to have been hitherto", that "the existing state of opinion as expressed in prices and the character of existing output is based on a correct summing up of future prospects", and "knowing that our own individual judgment is worthless, we endeavour to fall back on the judgment of the rest of the world which is perhaps better informed" (Keynes 1937). We project the present into the future, we comfort ourselves that our collective hunches have got the future right, we follow the herd.

Expectations formed in this way are bound to be fragile, since future circumstances almost inevitably turn out to be different from what we expected. Expectations will therefore be volatile, resulting in sudden shifts in investor (and speculator) sentiment. A shift in expectations can therefore suddenly change the values placed on assets, to the detriment of anyone whose assets are held in non-liquid form. As a consequence, money plays an essential role in a market economy because of its instant liquidity. The extent to which we desire to hold our wealth in the form of non-income-earning money rather than income-earning illiquid assets "is a barometer of the degree of our distrust of our own calculations and conventions concerning the future" (Keynes 1937).

This 'liquidity preference', Keynes argued, determines the rate of interest: the less we trust our fragile expectations of the future, the higher the rate of interest has to be to entice us to sacrifice unprofitable but safe cash for potentially profitable but volatile assets.

In assets themselves investors face two broad alternatives: lending money at the prevailing rate of interest (effectively purchasing bonds), or buying shares which confer part ownership of capital assets. Both these activities are effectively 'placement', as Blatt (1983) put it, rather than investment proper, however, which is the building of new capital assets.

New capital assets are produced not for their own sake, but in expectation of profits, and profits will come if their market prices (the result of placement activity) exceed their costs of construction. Physical investment is, therefore, extremely volatile because "it depends on two sets of judgments about the future, neither of which rests on an adequate or secure foundation – on the propensity to hoard [the flip side of liquidity preference] and on opinions of the future yield of capital-assets" (Keynes 1937). These two factors, which play a key role in determining how much investment takes place, are likely to feed upon and destabilise each other: if we become more pessimistic about the future prospects of investments, we are likely to want to hoard more, not less.

Having explained why uncertainty is so important in economic practice and economic theory, why expectations are fragile and volatile, and how these factors affect the rate of interest and the level of investment, Keynes returns once more to an attack on Say's Law. He divides expenditure into consumption – which is relatively stable – and investment – which is highly volatile – and emphasises that investment is the key determinant of the level and rate of change of output (and hence employment). His theory is therefore a theory "of why output and employment are so liable to fluctuation" (Keynes 1937).

Keynes concludes his summary as he began it, by arguing that conventional economics ignored the issue of uncertainty, and how expectations are formed under uncertainty, by simply assuming the problem away.

> I accuse the classical economic theory of being itself one of these pretty, polite techniques which tries to deal with the present by abstracting from the fact that we know very little about the future. (Keynes 1937)

Finally, in a departure from the General Theory of just a year earlier, Keynes criticises the concept of the 'marginal efficiency of capital' – the ratio of the yield of newly produced capital assets to their price. Whereas he used this concept extensively in the General Theory, here he argues that it is indeterminate since the price of capital assets is so volatile, and there will be a different 'marginal efficiency of capital' for every different level of asset

prices. Rather than being a determinant of investment, the 'marginal efficiency of capital' might simply be a by-product of the level of investment and current expectations.

These are all difficult concepts, especially for economists who were bred in the neoclassical tradition in which "at any given time facts and expectations were assumed to be given in a definite and calculable form; and risks, of which, though admitted, not much notice was taken, were supposed to be capable of an exact actuarial computation" (Keynes 1937).

But if Keynes had truly unleashed a revolution in economic thought, then economists should have attempted to get their minds around these difficult concepts, and fought to escape from the "habitual modes of thought and expression" which had gripped them prior to the calamity of the Great Depression.

Did economists do so? Some did, but the majority did not – and for that reason the profession bifurcated into two camps: a minority which swore fealty to Keynes's revolutionary vision, and a majority which paid lip service to some of Keynes's words, but which rapidly fell back into old, familiar ways. These economists ignored Keynes's General Theory, and even his pithy summary, instead clutching at another alleged summary by one J.R. Hicks.

Slimming Keynes down to size

Hicks's "Mr Keynes and the Classics' purported to be a book review of the General Theory. Hicks began by disputing that neoclassical economists held quite the views Keynes alleged that they held, and therefore tried to construct a more typical classical theory "in a form similar to that in which Mr. Keynes sets out his own theory . . . Thus I assume that I am dealing with a short period in which the quantity of physical equipment of all kinds available can be taken as fixed" (Hicks 1937).

Was this really the manner in which Keynes set out his own theory? Not according to Keynes, who criticised 'the classics' (by which he meant what we today call neoclassical economists) for working with a model in which "the amount of factors employed was given" (Keynes 1937: 212). Nonetheless, Hicks continued. He summarised the "typical Classical theory" in three equations, which argued that:

- The amount of money determined total output (output was some constant times the money stock);
- The rate of interest determined the level of investment; and
- The rate of interest determined the level of savings (and savings equalled investment).[75]

The first equation determines total output and total employment,[76] while the second two simply determine how much of output is devoted to investment, and how much to current consumption. If the savings rate increases, then so does investment. Increasing money wages "will necessarily diminish employment and raise real wages" (Hicks 1937), while the obverse policy – cutting money wages – will necessarily increase employment and reduce the real wage. Decreasing the money supply directly decreases income and employment, and is the main explanation for economic downturns (an argument which Milton Friedman later revived).

Clearly, Keynes's theory was substantially different from this. But how did Hicks summarise Keynes? In three more equations, where:

- The demand for money depends upon the rate of interest (in place of the 'Classical' fixed relationship between money and output);
- Investment is a function of the rate of interest; and
- Savings is a function of income.

Hello? What happened to uncertainty, expectations, liquidity preference determining the rate of interest, speculative capital asset prices? They are nowhere to be seen. Sometime later, Hyman Minsky commented that "Keynes without uncertainty is rather like Hamlet without the Prince" (Minsky 1975), but this is what Hicks served up as Keynes. Even the Readers' Digest would draw the line at this level of abridging, but this was not the end of Hicks's rephrasing of Keynes's Shakespearean sonnets into schoolyard doggerel.

He next argued that 'Keynes's' first equation omitted the impact of income on demand for money. This was the traditional 'transactions demand for money' argument that some level of money was needed to finance everyday transactions, so that an increase in income would generate an increase in the demand for money. To be truly general, said Hicks, the 'general theory' should include the impact of income on the demand for money, as well as the impact of the rate of interest.

Keynes had omitted discussion of the transactions demand for money because this demand was relatively stable, and therefore less important than the more important demand set by liquidity preference. But Hicks believed that "however much stress we lay upon the 'speculative motive', the 'transactions motive' must always come in as well" (Hicks 1937). So he proposed a revised set of equations in which the demand for money depends upon two variables – the rate of interest and the level of income –

though not, as Keynes had it, on "the degree of our distrust of our own calculations and conventions concerning the future" (Keynes 1937).

With this revision, Keynes, who was at such pains to distinguish himself from his predecessors – primarily though not exclusively on the basis of the importance he attached to uncertainty and expectations – is pushed back into the camp from which he desired to escape. As Hicks put it:

> With this revision, Mr. Keynes takes a big step back to Marshallian orthodoxy, and his theory becomes hard to distinguish from the revised and qualified Marshallian theories, which, as we have seen, are not new. Is there really any difference between them, or is the whole thing a sham fight? Let us have recourse to a diagram. (Hicks 1937)

Hicks's diagram explains why his rendition of Keynes was so readily accepted by economists, while Keynes's own summary was ignored. It was the old familiar totem of two intersecting curves, though now re-labelled to reflect its somewhat different derivation (see Figure 1).

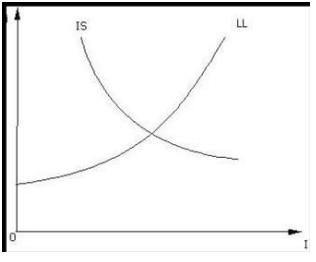


Figure 1: Hicks's model of Keynes

The downward-sloping curve, the equivalent of the microeconomic demand curve, was derived from the investment and savings relations in Hicks's model. The upward-sloping curve, the equivalent of the microeconomic supply curve, was derived from the money demand relation (on the assumption that the money supply was controlled by the monetary authorities, and was therefore determined outside the model).

The IS curve showed all those combinations of the rate of interest (i) and the level of income (I) which yielded equilibrium in the goods market. The LL curve (which economists today call the LM curve) showed all those combinations of the rate of interest and the level of income which gave equilibrium in the money market.

The IS curve was derived as shown in Figure 2. Investment was treated as a downward-sloping function of the rate of interest, so that a lower rate of interest would generate a higher level of investment. Investment determined the level of income via the mechanism Keynes called the multiplier.[77] Savings was an upward-sloping function of the level of income, so that a higher level of income would result in a higher level of savings. The IS curve shows all the combinations of the level of income and the rate of interest which made the level of savings compatible with the level of investment.

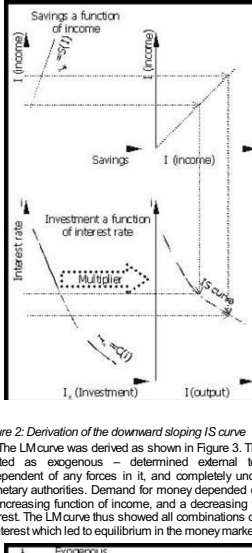


Figure 2: Derivation of the downward sloping IS curve

The LM curve was derived as shown in Figure 3. The money supply was treated as exogenous – determined external to the model itself, independent of any forces in it, and completely under the control of the monetary authorities. Demand for money depended on two factors: it was an increasing function of income, and a decreasing function of the rate of interest. The LM curve thus showed all combinations of income and the rate of interest which led to equilibrium in the money market.

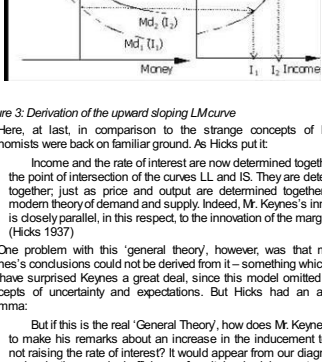


Figure 3: Derivation of the upward sloping LM curve

Here, at last, in comparison to the strange concepts of Keynes, economists were back on familiar ground. As Hicks put it:

> Income and the rate of interest are now determined together at P, the point of intersection of the curves LL and IS. They are determined together; just as price and output are determined together in the modern theory of demand and supply. Indeed, Mr. Keynes's innovation is closely parallel, in this respect, to the innovation of the marginalists. (Hicks 1937)

One problem with this 'general theory', however, was that many of Keynes's conclusions could not be derived from it – something which would not have surprised Keynes a great deal, since this model omitted his key concepts of uncertainty and expectations. But Hicks had an apparent dilemma:

> But if this is the real 'General Theory', how does Mr. Keynes come to make his remarks about an increase in the inducement to invest not raising the rate of interest? It would appear from our diagram that a rise in the marginal-efficiency-of-capital schedule must raise the curve IS; and, therefore, although it will raise income and employment, it will also raise the rate of interest. (Hicks 1937)

To Keynes, the reason why an increased desire to invest would not necessarily raise the rate of interest is because the latter was determined by liquidity preference, which "is a barometer of the degree of our distrust of our own calculations and conventions concerning the future" (Keynes 1937). In a depressed economy, an increase in investment could well reduce the "degree of distrust", leading to a fall in the rate of interest rather than a rise. But with Hicks's picture of Keynes shorn of uncertainty, conventions and

expectations, there were no such mechanisms to draw upon. Fortunately, Hicks's model provided a simple and far more conventional solution: simply bend the curves:

> This brings us to what, from many points of view, is the most important thing in Mr. Keynes's book. It is not only possible to show that a given supply of money determines a certain relation between Income and interest (which we have expressed by the curve LL); it is also possible to say something about the shape of the curve. It will probably tend to be nearly horizontal on the left, and nearly vertical on the right. This is because there is (1) some minimum below which the rate of interest is unlikely to go, and (though Mr. Keynes does not stress this) there is (2) a maximum to the level of income which can possibly be financed with a given amount of money. If we like we can think of the curve as approaching these limits asymptotically. (Hicks 1937)

This 'liquidity trap' enabled Hicks to provide an explanation for the Great Depression, and simultaneously reconcile Keynes with 'the Classics'. Keynes was consigned to one end of the LM curve where the liquidity trap applied, and 'the Classics' to the other where full employment was the rule (see Figure 4). In the 'classical' range of the LM curve, conventional economics reigned supreme: there was a maximal, full employment level of income, where any attempts to increase output would simply cause a rising interest rate (or inflation, in extensions of the IS-LM model). In the 'Keynesian' region, monetary policy (which moved the LM curve) was ineffective, because the LM curve was effectively horizontal, but fiscal policy (which moved the IS curve) could generate greater output – and hence employment – without increasing interest rates. A higher level of government expenditure could shift the IS curve to the right, thus moving the point of intersection of the IS and LM curves to the right and raising the equilibrium level of output.

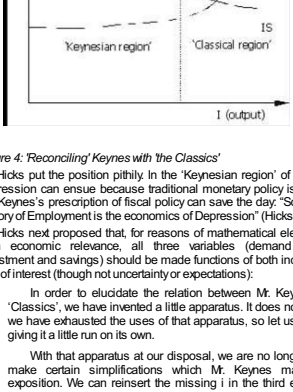


Figure 4: 'Reconciling' Keynes with 'the Classics'

Hicks put the position pithily. In the 'Keynesian region' of his model, a depression can ensue because traditional monetary policy is ineffective – but Keynes's prescription of fiscal policy can save the day: "So the General Theory of Employment is the economics of Depression" (Hicks 1937).

Hicks next proposed that, for reasons of mathematical elegance rather than economic relevance, all three variables (demand for money, investment and savings) should be made functions of both income and the rate of interest (though not uncertainty or expectations):

> In order to elucidate the relation between Mr. Keynes and the 'Classics', we have invented a little apparatus. It does not appear that we have exhausted the uses of that apparatus, so let us conclude by giving it a little run on its own.
>
> With that apparatus at our disposal, we are no longer obliged to make certain simplifications which Mr. Keynes makes in his exposition. We can reinsert the missing i in the third equation, and allow for any possible effect of the rate of interest upon saving; and, what is much more important, we can call in question the sole dependence of investment upon the rate of interest, which looks rather suspicious in the second equation. Mathematical elegance would suggest that we ought to have I and i in all three equations, if the theory is to be really General. (Hicks 1937)

Economists, having been threatened by Keynes with the need to completely retrain themselves, could now engage in their favourite game of tobogganing up and down one curve, moving another to the left or right, just as they did in microeconomics. It is little wonder that this Hicksian IS-LM model was adopted as the basis for 'Keynesian' economics, and equally little wonder that, many years later, macroeconomics was converted to a

subset of microeconomics.

Though 'cutting edge' economic analysis has left Hicks's model behind, most macroeconomists still think in IS-LM terms, and this model is still the common fare served up to undergraduate students as Keynesian economics. It therefore still has pedagogic and disciplinary relevance. So the question arises: from where did this model emanate? It clearly was not derived from Keynes's General Theory, apart from the adoption of some of Keynes's terminology. The mystery of its origins was finally solved by one Sir John Hicks – and older, benighted and somewhat wiser J.R. Hicks.

ISLM an apology

Hicks's detective work was published in a paper entitled 'IS-LM: An Explanation', but in many ways it was an apology. Published in the non-orthodox Journal of Post Keynesian Economics in 1980, the paper's opening sentence was:

> The IS-LM diagram, which is widely, though not universally, accepted as a convenient synopsis of Keynesian theory, is a thing for which I cannot deny that I have some responsibility. (Hicks 1980)

Even after this opening, Hicks clung to a very Walrasian vision of Keynes. But he conceded that his rendition had erroneously omitted any discussion of uncertainty or expectations. His explanation as to how he could have missed so fundamental an aspect of Keynes's thought was that, shortly before the General Theory was published, he had published a paper which, he believed, had strong similarities to Keynes's argument. What he then published as a review of Keynes was actually a restatement of his own model, using some of Keynes's terminology.

Hicks saw two key problems in cross-dressing as Keynes. The first was that his model was "a flexprice model . . . while in Keynes's the level of money wages (at least) was exogenously determined" (Hicks 1980); the second "more fundamental" problem was that Hicks's model used a period of a single week, while Keynes used "a 'short-period', a term with connotations derived from Marshall; we shall not go far wrong if we think of it as a year" (Hicks 1980).

Discussing the second problem, Hicks argued that the difference in period length had a drastic impact upon the relevance of expectations. With a time period of just a week, it is not unreasonable to keep expectations constant – and therefore to ignore them. But keeping expectations constant over a year in an IS-LM model does not make sense, because

> for the purpose of generating an LM curve, which is to represent liquidity preference, it will not do without amendment. For there is no sense in liquidity, unless expectations are uncertain. (Hicks 1980)

This was precisely the point Keynes himself made, in ironic form, in 1937:

> Money. . . . is a store of wealth. So we are told, without a smile on the face. But in the world of the classical economy, what an insane use to which to put it! For it is a recognized characteristic of money as a store of wealth that it is barren; whereas practically every other form of storing wealth yields some interest or profit. Why should anyone outside a lunatic asylum wish to use money as a store of wealth?
>
> Because, partly on reasonable and partly on instinctive grounds, our desire to hold Money as a store of wealth is a barometer of the degree of our distrust of our own calculations and conventions concerning the future . . .The possession of actual money lulls our disquietude; and the premium which we require to make us part with money is the measure of the degree of our disquietude. (Keynes 1937)

Thus, without uncertain expectations, there is no sense in liquidity preference either, and Hicks cannot justify the LM half of his IS-LM model. But with uncertain expectations, there is no sense in equilibrium analysis either, since equilibrium can only be maintained if expectations are continually being fulfilled (and are therefore irrelevant). Hicks concluded that the equilibrium/constant expectations framework of the IS-LM model was theoretically unsound, and practically irrelevant to the problems of the macro economy:

> I accordingly conclude that the only way in which IS-LM analysis usefully survives – as anything more than a classroom gadget, to be superseded, later on, by something better – is in application to a particular kind of causal analysis, where the use of equilibrium methods, even a drastic use of equilibrium methods, is not inappropriate . . .
>
> When one turns to questions of policy, looking towards the future instead of the past, the use of equilibrium methods is still more suspect. For one cannot prescribe policy without considering at least the possibility that policy may be changed. There can be no change of policy if everything is to go on as expected – if the economy is to remain in what (however approximately) may be regarded as its existing equilibrium. (Hicks 1980)

From IS-LM to the representative agent

Hicks's revelation came far too late to stop the economic juggernaut he had set in motion – his contrary position was generally ignored by economists, who were more inclined to put his views down to approaching senility than to any blinding logical revelation. In any case, the gradual demolition of IS-LM by economists was substantially advanced by 1980.

This demolition began back in the 1950s with the critique that Hicks's 'Keynesian' model did not have good microeconomic foundations: it was not possible to derive the kind of results given by the model – such as the economy settling into a less than full-employment equilibrium – from standard microeconomics.

Keynes himself would have had no problem with this, since he tried to distinguish himself from the 'Classical' theory in which the presumption of certainty played such a central role. In an uncertain world, none of the old-fashioned notions of self-correcting tendencies would operate. It would be better to derive a new microeconomics for a world of uncertainty than to remake macroeconomics in microeconomic's image.

But of course, there was no uncertainty in the emasculated version of Keynes which survived Walras' Law on the one hand, and Hicks's IS-LM 'summary' on the other. Economists who accepted this melange as Keynes had no inkling that uncertainty – or any other similarly unfamiliar concept – played any role in Keynes's economics. To them, the critique that Keynesian macroeconomics lacked sound microeconomic foundations was a telling one. What Keynes's 'true believers' called the 'Keynesian counter-revolution' was under way.

There were many steps along this path, only two of which warrant discussion today: the 'rational expectations' and 'representative agent' approaches to macroeconomics.

Rational expectations

The basis for the rational expectations approach to economics was, ironically enough, that Keynesians had not paid sufficient attention to the role of expectations in macroeconomics.

It should be evident that anyone who believed this had either never read Keynes, or had not comprehended what they did read. However, having at least apparently started down a track which Keynes would have supported – to introduce a formal analysis of expectations into economics – the proponents of this school did something quite contrary.

> *They presumed that the expectations of economic agents would be the same as the predictions of conventional IS-LM macroeconomic model, and they presumed that these predictions were accurate.*

One of these predictions was that increasing the money supply would cause inflation. In the model without expectations, if the government increased the money supply in order to reduce unemployment, there would be a lag between when the money supply was increased, and when the inflation actually occurred. In the meantime, the increased money supply would have the impact desired by the government, of increasing economic activity – and hence reducing unemployment.

The twist of adding expectations into the model, when expectations were identical to the prediction of the model, was that inflation would occur instantly, rather than with a lag. This is because, since everyone expected an increased money supply to cause inflation, everyone instantly put their prices up as soon as the money supply rose. The lag between an increase in the money supply and an increase in prices was eliminated, and with it went any temporary impact of the money supply on unemployment.

Not surprisingly, this doctrine was termed the 'policy ineffectiveness proposition'. If anything that was consciously done by policy makers to manipulate the economy led instantly to countervailing behaviour by people in the economy, then nothing the government could do would alter the rate of unemployment. Instead, all the government could do would be cause inflation.

This theory was very modish for quite a while, and won its originators the Nobel Prize in economics. However, it has gradually receded into obscurity as its empirical predictions were manifestly contradicted by the data – though it still forms a large part of the modern training of an undergraduate economist.

In the end, the main role of rational expectations macroeconomics was to completely discredit Hicks's IS-LM concepts, which had for so long formed the core of macroeconomic thinking (though they too continue to be fed to undergraduates). The process which Hicks himself had begun many years earlier concluded with macroeconomics being completely gutted. Into the vacuum stepped microeconomics, in the guise of the 'representative agent' model of the macroeconomy.

Representative agent macroeconomics

You will remember from Chapter 2 that the representative agent was a kludge invented by Gorman to get around the problem that, in general, the preferences of individual agents could not be aggregated. I have more to

say about this in Chapter 12, but here it is worth noting that representative agent macroeconomics amounts to assuming that the economy consists of a single individual producing and consuming a single commodity. However complex might be the reasoning used by such aficionados as Paul Krugman, the realm of applicability of this theory is thus that of Robinson Crusoe, living off coconuts before the arrival of Man Friday. It beggars belief that anyone who truly knew where this notion had come from would attempt to apply it to something as complex as a modern, multi-commodity, multi-million person economy.

So What?

Macroeconomics is economic policy par excellence, but economic theory itself has virtually reached the position that there should be no macroeconomic policy. The clearest evidence of this is the Maastricht Treaty, which made restricting budget deficits to no more than 3% of GDP a condition for membership of the European Union. While some fudging has been allowed to make membership possible in the first place, when an economic crisis eventually strikes, Europe's governments may be compelled to impose austerity upon economies which will be in desperate need of a stimulus.

The belief that a capitalist economy is inherently stabilising is also one for which inhabitants of market economies may pay dearly in the future. As they were initially during the Great Depression, economists today may be the main force preventing the introduction of countervailing measures to any future economic slump. Economics may make our recessions deeper, longer and more intractable, when the public is entitled to expect economics to have precisely the opposite effect.

A reprieve?

Fortunately for economists, the macroeconomy – at least in the United States – appeared to be functioning fairly well at the end of the year 2000. It is thus possible for economists to believe and preach almost anything, because they can bask in the entirely coincidental fact that the macroeconomy appears healthy.

However, this accidental success may not last long if the pressures which have been clearly growing in the financial side of the economy finally erupt.

10. The price is not right

Why finance markets can get the price of assets so badly wrong

The Internet stock market boom was the biggest speculative bubble in world history.

Other manias have involved more ridiculously overvalued assets, or more preposterous objects of speculation – such as the tulip craze in 17th century Holland, the South Sea Bubble and Mssissippi Bubble of the 18th century, or Japan's 1980s Bubble Economy speculation over Tokyo real estate. But no other bubble – not even the 'Roaring Twenties' boom prior to the Great Depression – has involved so many people, speculating so much money, in so short a time, to such ridiculous valuations.

But of course, an economist wouldn't tell you that. Instead, economists have assured the world that the stock market's valuations reflect the true future prospects of companies. The most famous – and fatuous – such assurance is given in Dow 36,000, which its authors were defending even when the Dow had officially entered a correction[78] from its all-time high of March 2000, and the Nasdaq was firmly in bear market territory (Time, 22 May 2000: 92–93).[79] The mammoth valuations, argued Hassett and Glassman, were simply the product of investors reassessing the risk premiums attached to stocks, having realised that over the long term, stocks were no riskier than bonds.

Economists were similarly reassuring back in 1929, with the most famous such utterance being Irving Fisher's comment that:

> Stock prices have reached what looks like a permanently high plateau. I do not feel that there will soon, if ever, be a fifty or sixty point break below present levels, such as Mr. Babson has predicted. I expect to see the stock market a good deal higher than it is today within a few months. (Irving Fisher, 15 October 1929)

This was published less than two weeks before 'Black Monday', 28 October 1929, when the Dow Jones Industrial Average closed 12.8% below its previous level, and fell another 11.7% the following day. In just 15 days of wild gyrations from the day of Fisher's comments, the market fell over 120 points (from a level of about 350): twice as far as even Fisher's bearish rival Babson had predicted, and twice as much as Fisher had believed would ever be possible. Three years later, the stock market indices had fallen 90%, and many a once-rich speculator was bankrupt. Investors who trusted economists back then lost their shirts. Trusting souls who accept economic assurances that markets are efficient are unlikely to fare any better this time when the Bull gives way to the Bear.

The kernel

> "There's glory for you!"
>
> "I don't know what you mean by 'glory'," Alice said.
>
> Humpty Dumpty smiled contemptuously. "Of course you don't – till I tell you. I meant 'there's a nice knock-down argument for you!'"
>
> "But 'glory' doesn't mean 'a nice knock-down argument'," Alice objected.
>
> "When I use a word," Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean – neither more nor less."

All sciences invent their own language, just as Lewis Carroll's famous egghead invented his own meanings for words. Many sciences harness words which in are common usage, but give them a quite different technical meaning. But no other science plays so fast and loose with the English language as economics.

Physics, for example, calls the fundamental constituents of matter 'strings'. This isn't particularly confusing, since it's obvious that physicists don't believe that a length of yarn is the basic unit of matter.

However, when economists call stock markets 'efficient', the usage is nowhere near as clear cut. A colloquial meaning of efficient is 'does things quickly with a minimum of waste', and it's clear that this meaning can apply to modern, computerised, Internet-accessible bourses. Thus it often seems reasonable to the public that economists describe finance markets as 'efficient'.

However, when economists say that the stock market is efficient, they mean that they believe that stock markets accurately price stocks on the basis of their unknown future earnings. That meaning shifts 'efficient' from something which is obvious to something which is a debatable proposition. But that's not the end of the story, because to 'prove' that markets are efficient in this sense, economists make three bizarre assumptions:

- That all investors have identical expectations about the future prospects of all companies;
- That these identical expectations are correct; and

Ÿ That all investors have equal access to unlimited credit.

Clearly, the only way these assumptions could hold would be if each and every stock market investor were God. Since in reality the stock market is inhabited by mere mortals, there is no way that the stock market can be efficient in the way that economists define the term. Yet economists assert that stock markets are 'efficient', and dismiss criticism of these assumptions with the proposition that you can't judge a theory by its assumptions. As Chapter 7 showed, this defence is bunk.

In a way, it's fitting that Lewis Carroll put those words in Humpty Dumpty's mouth, rather than equally appropriate vessels such as the Mad Hatter, or the Red Queen. Humpty Dumpty, after all, had a great fall…

The roadmap

The chapter begins by considering the development over time of the prevailing attitude to finance, starting with the medieval prohibition against the lending of money at interest, and culminating in economists treating the lending of money as no different from any other commodity exchange. The main economist responsible for the economic theory of lending was Irving Fisher, who, as just mentioned, effectively went bankrupt during the Depression by following his own theories. However, he subsequently developed a quite different theory, which argued that excessive debt and falling prices could cause depressions. After outlining this theory, I consider the modern theory of finance known as the 'efficient markets hypothesis'. The validity of the assumptions needed to buttress this theory is assessed in the light of the logic outlined in Chapter 7. Since these are domain assumptions, the theory is inapplicable in the real world, so that markets cannot possibly be 'efficient' as economists define the term.

Fisher on finance: from reassuring oracle to ignored Cassandra

Irving Fisher was one of the many victims of the Great Crash of 1929, losing a fortune worth over $100 million in today's dollars, and being reduced to penury.[80] But his greater loss, in many ways, was one of prestige. Before this infamous utterance, he was easily America's most respected and famous economist, renowned for developing a theory of money that explained the valuation of financial assets. After it, he was a pariah.

This was a great pity, because in the depths of the Great Depression, he developed an explanation of how financial speculation could lead to economic collapse. However, this new theory – which rejected many of the assumptions of his previous model of finance – was ignored. Instead, Fisher's pre-Great Depression theory of finance continued as the economic theory of how asset prices are determined.

Decades later, Fisher's 'Debt Deflation Theory of Great Depressions' was rediscovered by the non-orthodox economist Hyman Minsky, while at much the same time Fisher's pre-Great Depression theory was formalised into the efficient markets hypothesis. Fisher thus has the dubious distinction of fathering both the conventional theory of finance – which, like his 1929 self, reassures finance markets that they are rational – and an unconventional theory which argues that speculative bubbles can cause economic depressions.

Pre-Depression Fisher: the time value of money

In 1930 Fisher published The Theory of Interest, which asserted that the interest rate "expresses a price in the exchange between present and future goods" (Fisher 1930).[81] This argument was a simple extension of the economic theory of prices to the puzzle of how interest rates are set, but it has an even older genealogy: it was first argued by Jeremy Bentham, the true father of modern neoclassical economics, when in 1787 he wrote 'In Defence of Usury'.

Never a lender nor a borrower be…

Today, usury means lending money at an exorbitant rate of interest; in antiquity, it meant lending money at any rate of interest at all. However, the medieval objection was not to the rate of interest itself, but to the lender's desire to profit from a venture without sharing in its risks. A usurious contract was one in which the lender was guaranteed a positive return, regardless of whether the borrower's venture succeeded or failed:

> The primary test for usury was whether or not the lender had contracted to lend at interest without assuming a share of the risk inherent to the transaction. If the lender could collect interest regardless of the debtor's fortunes he was a usurer. (Jones 1989)

As trade came to play a larger role in society, the prohibitions against usury were weakened, and the legal definition was modified to match today's colloquial meaning.[82] By Bentham's time, the legal definition

referred to a rate of interest greater than 5%.

Adam Smith supported this legal limit. Smith argued that the complete prohibition, "like all others of the same kind, is said to have produced no effect, and probably rather increased than diminished the evil of usury" (Smith 1776). However, Smith supported the concept of a legal limit to the rate of interest set just above the going market rate.[83] because such a limit actually improved the allocation of the country's credit. The advantage of a legal limit, according to Smith, was that when set properly it excluded only loans to 'prodigals and projectors', thus making more of the country's capital available for loan to industrious people:

> The legal rate . . . ought not to be much above the lowest market rate. If the legal rate of interest in Great Britain, for example, was fixed so high as eight or ten per cent, the greater part of the money which was to be lent would be lent to prodigals and projectors, who alone would be willing to give this high interest . . . Agreat part of the capital of the country would thus be kept out of the hands which were most likely to make a profitable and advantageous use of it, and thrown into those which were most likely to waste and destroy it. Where the legal rate of interest, on the contrary, is fixed but a very little above the lowest market rate, sober people are universally preferred, as borrowers, to prodigals and projectors. The person who lends money gets nearly as much interest from the former as he dares to take from the latter, and his money is much safer in the hands of the one set of people than in those of the other. A great part of the capital of the country is thus thrown into the hands in which it is most likely to be employed with advantage. (Smith 1776)

In defence of usury

Bentham's rejoinder to Smith's arguments may well have set the standard for fanciful and specious reasoning to which modern economics has since aspired.

Smith referred to two classes of borrowers who could be expected to accede to rates of interest substantially above the lowest market rate: 'prodigals and projectors'. The former are individuals who would waste the money on conspicuous consumption; the latter are those who promote ill-conceived schemes to the public, which result in inappropriate investment. Smith's case in favour of a legal ceiling to the rate of interest thus had both a 'microeconomic' and a 'macroeconomic' aspect.

Macroeconomics was Smith's key concern: encouraging "prodigals and projectors" would result in "a great part of the capital of the country" being thrown into the hands of "those which were most likely to waste and destroy it". The ceiling, by removing the incentive to lend to such borrowers, would result in a higher overall quality of investment, and thus higher growth.

Bentham's riposte ignored macroeconomics. Instead, it began from the microeconomic and libertarian presumption that "no man of ripe years and of sound mind, acting freely, and with his eyes open, ought to be hindered, with a view to his advantage, from making such bargain, in the way of obtaining money, as he thinks fit" (Bentham 1787).

He initially conceded that the restraint of prodigal behaviour may give grounds for setting a ceiling to the rate of interest, only to then argue that in practice a prodigal would not be charged an exorbitant rate of interest. He began with the proposition that "no man . . . ever thinks of borrowing money to spend, so long as he has ready money of his own, or effects which he can turn into ready money without loss". Secondly, the exceptions to the above rule who have the requisite collateral can get a loan at the usual rate. Thirdly, those who do not have security will only be lent to by those who like them, and these friendly persons will naturally offer them the standard rate:

> Persons who either feel, or find reasons for pretending to feel, a friendship for the borrower, can not take of him more than the ordinary rate of interest: persons who have no such motive for lending him, will not lend him at all. (Bentham 1787)

If Bentham were to be believed, the friendly bank manager of the 1950s had many a precursor in 18th century Britain, while the rapacious Shylock perished with Shakespeare in the 16th.

A bit of empirical research would have revealed that, though rates of interest had fallen dramatically as finance became institutionalised, there was no shortage of lenders willing to hand prodigals ready cash at high rates of interest, in return for ownership of their assets should they go bankrupt. But Bentham's more important sleight of mind was to ignore the macroeconomic argument that the legislative ceiling to the rate of interest improved the overall quality of investment by favouring "sober people" over "prodigals and projectors".

The historical record favoured Smith. The 17th, 18th and 19th centuries are awash with examples of projectors promoting fantastic schemes to a gullible public. The most famous have entered the folklore of society: the Tulip Mania, the South Sea Bubble, the Mississippi Land Scheme (Mackay 1841). What has not sunk in so deeply is that the financial panics that occurred when these bubbles burst frequently ruined whole countries.[84]

However, the tide of social change and the development of economic theory favoured Bentham. The statutes setting maximum rates were eventually repealed, the concept of usury itself came to be regarded as one of those quaint preoccupations of a more religious age, and modern economics extended Bentham's concept that "putting money out at interest, is exchanging present money for future" (Bentham 1787). Of course, the magnificent edifice economists built upon Bentham's morsel assumed that everything happened in equilibrium.

The time value of goods

In keeping with the economic belief that the economy is fundamentally a barter system, in which money is merely a lubricant, Fisher restated Bentham's concept in terms of goods, rather than money: the rate of interest "expresses a price in the exchange between present and future goods" (Fisher 1930).

Fisher's model had three components: the subjective preferences of different individuals between consuming more now by borrowing, or consuming more in the future by forgoing consumption now and lending instead; the objective possibilities for investment; and a market which reconciled the two.

From the subjective perspective, a lender of money is someone who, compared to the prevailing rate of interest, has a low time preference for present over future goods. Someone who would be willing to forgo $100 worth of consumption today in return for $103 worth of consumption next year has a rate of time preference of 3%. If the prevailing interest rate is in fact 6%, then by lending out $100 today, this person enables herself to consume $106 worth of commodities next year, and has clearly made a personal gain. This person will therefore be a lender when the interest rate is 6%.

Conversely, a borrower is someone who has a high time preference for present goods over future goods. Someone who would require $110 next year in order to be tempted to forego consuming $100 today would decide that, at a rate of interest of 6%, it was worth his while to borrow. That way, he can finance $100 worth of consumption today, at a cost of only $106 worth of consumption next year. This person will be a borrower at an interest rate of 6%.

The act of borrowing is thus a means by which those with a high preference for present goods acquire the funds they need now, at the expense of some of their later income.

Individual preferences themselves depend in part upon the income flow that an individual anticipates, so that a wealthy individual, or someone who expects income to fall in the future, is likely to be a lender, whereas a poor individual, or one who expects income to rise in the future, is likely to be a borrower (Fisher 1930).

At a very low rate of interest, even people who have a very low time preference are unlikely to lend money, since the return from lending would be below their rate of time preference. At a very high rate of interest, even those who have a high time preference are likely to be lenders instead, since the high rate of interest would exceed their rate of time preference. This relationship between the rate of interest and the supply of funds gives us an upward-sloping supply curve for money.

The objective perspective reflects the possibilities for profitable investment. At a high rate of interest, only a small number of investment projects will be expected to turn a profit, and therefore investment will be low. At a low rate of interest, almost all projects are likely to turn a profit over financing costs, so the demand for money will be very high. This relationship between the interest rate and the demand for money gives us a downward-sloping demand curve for money.

The market mechanism brings these two forces into harmony by yielding the equilibrium rate of interest. This can, of course, be represented by the classic economic totem, as shown in Figure 1.

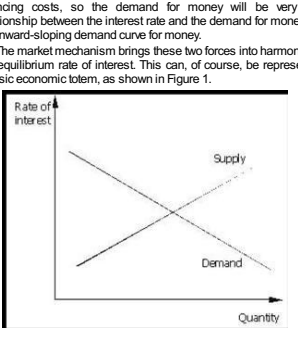


Figure 1:

Economics, it appears, is back in familiar territory. But there are some special, time-based nuances to the credit market. In the goods market, transactions occur immediately: one bundle of goods today is exchanged for another bundle of goods today. However, in the credit market, the 'purchaser' (the company offering an investment opportunity) takes immediate delivery of the loan, but repays principal and interest in instalments over time. Ancillary assumptions were therefore required to stretch the standard static vision of the market to the time-based creature that credit really is. These additional assumptions, in Fisher's words, were "(A) The market must be cleared – and cleared with respect to every interval of time. (B) The debts must be paid" (Fisher 1930).

Fisher saw nothing wrong with these ancillary assumptions, until he and countless others personally violated them during the Great Depression.

Fisher during the Crash: 'don't panic'

To his credit, Fisher's response to the Great Depression was worthy of Keynes's apocryphal statement that "when the facts prove me wrong, I change my mind". But at first Fisher clung to his pre-Crash optimism that the American economy was fundamentally sound, that a wave of invention had introduced a new era of higher productivity, that the new media of radio would revolutionise business. It all sounds so familiar today…

A new era…

Fisher's comments to a bankers' forum on 'Black Wednesday' – 23 October, when stocks fell by an unprecedented 6.3% in one day – confirm the old adage that 'the more things change, the more they remain the same'. Every factor that Fisher then thought justified the stock market's bull run has its counterpart today: it was 'a new era', a wave of invention (read 'the Internet') justified high valuations, stable prices reduced the uncertainty of share ownership, stocks were better long-term investments than bonds, investment trusts (read 'mutual funds') enabled much more intelligent stock selection, a debt-financed consumer boom was natural when a great increase in income was rationally anticipated.

Fisher first recounted the ways in which the 1929 stock market boom was remarkable. Shares had doubled in value since 1926, and any investor who had 'followed the herd' and bought and sold shares simply on the basis of their popularity would have increased her wealth tenfold in those three years. Stock prices had risen so much that dividend yields were below bond yields. Brokers' loans – effectively margin call lending – were at their highest level in history. All these observations supported the notion that the market "seems too high and to warrant a major bear movement" (Fisher 1929).

However, he then gave four reasons why the 1929 valuations were sensible: changed expectations of future earnings, reinvestment of dividends, a change in risk premiums, and a change in the way in which future income is discounted.

He supported the first argument with the statement that

> We are now applying science and invention to industry as we never applied it before. We are living in a new era, and it is of the utmost importance for every businessman and every banker to understand this new era and its implications… All the resources of modern scientific chemistry, metallurgy, electricity, are being utilized – for what? To make big incomes for the people of the United States in the future, to add to the dividends of corporations which are handling these new inventions, and necessarily, therefore, to raise the prices of stocks which represent shares in these new inventions. (Fisher 1929; emphasis added)

This wave of invention, with its return in years yet to come, meant that it was quite natural for the ratio of share price to historic earnings to rise. In fact, these new firms should be expected to make losses as they established their new inventions:

> In the airline industry very little attention is paid to the earnings today, because the price of the stock is purely a speculation on the far bigger returns that are expected in the future. Any new invention . . . at first does not give any profit… (Fisher 1929)

Low inflation also played a role in high stock valuations, since a stable price level gives "an immense impulse towards prosperity" (Fisher 1929).

The second factor, the reinvestment of dividends, was a positive force since firms that did this – rather than handing dividends back to investors – were able to grow more rapidly. Hence "many of the stocks which sell the highest on the stock exchange and which have had the most spectacular rise are not paying any dividends" (Fisher 1929).

The third reason, a change in the way the public estimates risk, occurred because Edgar Smith's influential book Common Stocks as Long Term Investments had shown that over the longer term, stocks outperformed bonds. As a result, "[t]here has been almost a stampede towards stocks, and away from bonds" (Fisher 1929; today, Hassett and Glassman's Dow 36,000 and its ilk spread the same delusion).

This movement had led to the establishment of the new profession of investment counselling, and then the new institution of investment trusts which "can afford to make studies of stocks which the individual investor could not study" (Fisher 1929). As well as diversifying and spreading risk, these institutions enabled stocks to be scientifically selected. This explained why "our stock market is highly selective today", and as a result Fisher wasn't troubled by the fact that:

> Half of the stocks during the last year have fallen in spite of the fact that the average as shown by the index numbers had risen. The leaders are becoming fewer and fewer, and those stocks that are leading have a greater and greater scarcity value. (Fisher 1929)

Fisher conceded that rank speculation played some role in the market, but he blamed this "lunatic fringe" more for the crash in stock prices than for its run-up over the preceding four years:

> There is a certain lunatic fringe in the stock market, and there always will be whenever there is any successful bear movement going on . . . they will put the stocks up above what they should be and, when frightened, . . . will immediately want to sell out. (Fisher 1929)

This speculative fringe ranked 15th out of Fisher's 15 determinants of the level of stock prices, though he was not so confident of his ranking after the market's 6% fall on Black Wednesday. Nonetheless, he still argued that "the other fourteen causes are far more important than this one cause itself". He acknowledged that most speculation took place with borrowed money – a theme that would later become his bte noir. But he argued that most of this money had been borrowed to finance consumption today – rather than just rank speculation – because consumers were simply cashing in on rationally anticipated future increases in income:

> To a certain extent it is normal that during an era such as we are now passing through, where the income of the people of the United States is bound to increase faster perhaps than ever before in its history, and it has during the last few years increased amazingly, that we should try to cash in on future income in advance of its occurring, exactly on the principle that when a young man knows he has been given unexpectedly a large bequest, and that it will be in his hands inside a year, he will borrow against it in advance. In other words, there ought to be a big demand for loans at a high rate of interest during a period of great increase in income. (Fisher 1929)

He concluded with an expectation that the market's 12% fall in the preceding eight days was an aberration:

> Great prosperity at present and greater prosperity in view in the future . . . rather than speculation . . . explain the high stock markets, and when it is finally rid of the lunatic fringe, the stock market will never go back to 50 per cent of its present level . . . We shall not see very much further, if any, recession in the stock market, but rather . . . a resumption of the bull market, not as rapidly as it has been in the past, but still a bull rather than a bear movement. (Fisher 1929)

Fisher after the crash: the debt-deflation hypothesis

Fisher was, of course, profoundly wrong, and at great personal cost to himself. The market receded 90% from its peak, and the index did not regain its 1929 level for a quarter of a century.[85] As the crash persisted, the slump deepened into the Great Depression, with, at its nadir, over 25% of America's workers unemployed. Fisher's personal fortune evaporated, and his perspective on the American financial system shifted from one of confidence to one of alarm.

He eventually developed a radically different analysis of finance, one in which his ancillary assumptions in The Rate of Interest – that "the market must be cleared, and cleared with respect to every interval of time" and that "The debts must be paid" – were systematically violated. Now he acknowledged that the market was never in equilibrium, and that debts could fail to be repaid, not just individually but en masse. Static reasoning gave way to an analysis of the dynamic forces which could have caused the Great Depression.

Whereas he had previously assumed that the economy was always in equilibrium, now he appreciated that even if the real economy actually momentarily reached equilibrium, this state would be short-lived since "new disturbances are, humanly speaking, sure to occur, so that, in actual fact, any variable is almost always above or below the ideal equilibrium" (Fisher 1933).

Equilibrium was also likely to be precarious. Whereas beforehand he had simply taken it for granted that equilibrium was stable, now he realised that equilibrium "though stable, is so delicately poised that, after departure from it beyond certain limits, instability ensues". A slight movement away from equilibrium could set in train forces that would drive the economy even further away, rather than returning it to balance.

While any of a multitude of factors could, according to Fisher, push the system away from equilibrium, the crucial ingredient needed to turn this

limited instability into a catastrophic collapse was an excessive level of debt, where "the breaking of many debtors constitutes a 'crash', after which there is no coming back to the original equilibrium".

He ventured the opinion that the "two dominant factors" that cause depressions are "over-indebtedness to start with and deflation following soon after". Though other factors are important, debt – the entry into a contractual obligation to repay principal with interest – and a falling price level are crucial:

> Thus over-investment and over-speculation are often important; but they would have far less serious results were they not conducted with borrowed money. That is, over-indebtedness may lend importance to over-investment or to over-speculation. The same is true as to over-confidence. I fancy that over-confidence seldom does any great harm except when, as, and if, it beguiles its victims into debt. (Fisher 1933)

The final sentence in this quote is rather poignant, since Fisher himself was a classic instance of someone whom over-confidence had beguiled into debt.[86]

Overconfidence leads investors to overestimate the prospective gain from investment, or to underestimate the risks, and thus commit themselves to an unsustainable level of debt. In either case, the investor commits funds well beyond the level which returns an optimum gain. Such overconfidence is an inevitability in the real world because, as noted above, all real-world variables are bound to be either above or below their ideal equilibrium values.

A chain reaction then ensues that can tip the economy into depression. It begins with distress selling, at severely reduced prices, driven by the need to cover debt repayments. Falling prices means that the real burden of debt actually rises, even as nominal debt is reduced, and the repayment of debts also reduces the money supply. These effects cause further bankruptcies, reducing profits, investment, output and employment. Pessimism rises, causing those with money to hoard it, which further reduces business activity. The falling price level also has the perverse effect that the real rate of interest rises even though nominal rates have fallen, and this drastically reduces investment (Fisher 1933).

Fisher's theory was thus an alternative explanation of the Great Depression to both Keynes's rejection of Say's Law and Hicks's 'liquidity trap' (discussed in Chapter 9). But though the chain reaction argument is plausible, Fisher provided no formal proof for it – in contrast to his previous emphasis upon formal mathematical reasoning. Partly for this reason, his thesis was received poorly by the economics profession, and his insights were swamped by the rapid adoption of Hicks's IS-LM analysis after the publication of Keynes's General Theory.[87]

After the Great Depression, economists continued to cite his pre-crash work on finance, while his debt-deflation theory was largely ignored.[88] As a result, the antipathy he saw between the formal concept of equilibrium and the actual performance of asset markets was also ignored. Equilibrium once again became the defining feature of the economic analysis of finance. This process reached its zenith with the development of what is known as the 'efficient markets hypothesis'.

The efficient markets hypothesis

Non-economists often surmise that the term 'efficient' refers to the speed at which operations take place on the stock market, and/or the cost per transaction. Since the former has risen and the latter fallen dramatically with computers, the proposition that the stock market is efficient appears sensible. Market efficiency is often alleged to mean that 'investors are assumed to make efficient use of all available information', which also seems quite reasonable.

However, the economic concept of efficiency means something quite different from the normal parlance. In the case of the stock market, it means at least four things:

Ÿ That the collective expectations of stock market investors are accurate predictions of the future prospects of companies;

Ÿ That share prices fully reflect all information pertinent to the future prospects of traded companies;

Ÿ That changes in share prices are entirely due to changes in information relevant to future prospects, where that information arrives in an unpredictable and random fashion; and

Ÿ That therefore stock prices 'follow a random walk', so that past movements in prices give no information about what future movements will be – just as past rolls of a dice can't be used to predict what the next roll will be.

These proposition are a collage of the assumptions and conclusions of the 'efficient markets hypothesis' ('EMH') and the 'capital assets pricing model' ('CAPM'), which were formal extensions to Fisher's (pre-Depression) time value of money theories. Like the Fisher theories of old,

these new theories were microeconomic in nature, and presumed that finance markets are continually in equilibrium. There were several economists who developed this sophisticated equilibrium analysis of finance. In what follows I will focus on the work of W.F. Sharpe.

Risk and return

It seems reasonable to argue that an asset that gives a high return is likely to be riskier than one that gives a lower return. If an investor wants complete safety then she can invest in government bonds. If a higher rate of return is desired, then she can invest in corporate bonds, or shares. The former hold the risk of default, while the latter can rise or fall unpredictably in price, and do not have a guaranteed income flow.

In equilibrium in an asset market, this would result in the situation depicted in Figure 2.

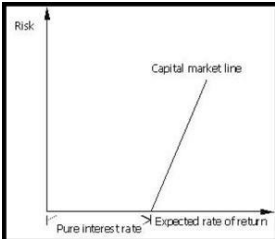


Figure 2:

Sharpe provided an explanation for this equilibrium position in terms of the theory of individual behaviour discussed in Chapter 2. Once again, we find ourselves tobogganing up and down indifference curves.

The individual rational investor

Sharpe began by assuming that "an individual views the outcome of any investment in probabilistic terms; he is willing to act on the basis of . . . expected value and standard deviation" (Sharpe 1964 [1991]).[89] An investor gets greater utility from a higher return than a lower one, and lower utility from an asset with a high standard deviation than a lower one. This assumption enabled Sharpe to plot an investor's preferences in terms of indifference curves.

However, there was one twist compared to standard indifference curve analysis as outlined in Chapter 2. Risk is a 'bad', not a 'good' – and a consumer maximises his utility by experiencing as little risk as possible. Therefore, rather than being shaped like a slippery dip, these indifference curves are shaped like a rainbow: they curve upwards to the right, rather than downwards. The direction of higher utility is now down and to the right, rather than up and to the right.

An investor's utility rose as expected returns rose and the variability of returns fell. This is shown by curves I, II and III in Figure 3, with higher utility occurring as you move right (higher expected return) and down (lower expected risk).

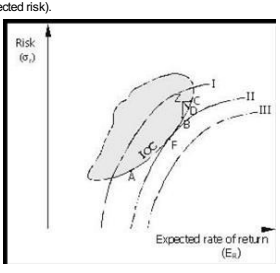


Figure 3:

Sharpe then argued that all possible investments could also be described in terms of their expected returns and the standard deviation of

their expected returns. These could have any pattern at all – some investments would have a very high expected return and low variability, others a low expected return and high variability, and so on. This is represented as a shaded 'cloud' of possible investments in Figure 3, the edge of which happens to be tangential to indifference curve II at the position F.

With this picture of investor behaviour, Sharpe can show that the only investments that are rational for this investor are those that fall on the edge of the cloud of possible investments, which he labels the 'investment opportunity cloud' or IOC (Sharpe 1964 [1991]).12 If this were the end of the matter, then this individual investor would invest in F, and that would be that.

However, all investments along the IOC have some risk. Sharpe assumes the existence of some asset which has no risk, and that therefore is located on the horizontal axis (where the standard deviation is zero). Our investor in Figure 3 could put part of her portfolio into F, and part into this riskless asset (P in Figure 4). Sharpe also assumes that individuals can borrow or lend as much money as they like at this riskless interest rate.

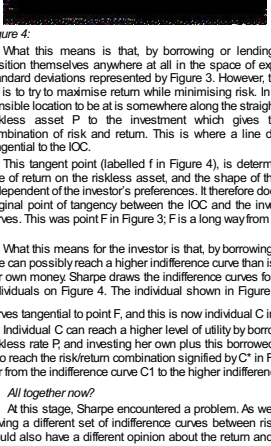


Figure 4:

What this means is that, by borrowing or lending, an investor can position themselves anywhere at all in the space of expected returns and standard deviations represented by Figure 3. However, the rational thing to do is to try to maximise return while minimising risk. In that case, the only sensible location to be at is somewhere along the straight line that links this riskless asset P to the investment which gives the best possible combination of risk and return. This is where a line drawn through P is tangential to the IOC.

This tangent point (labelled f in Figure 4), is determined simply by the rate of return on the riskless asset, and the shape of the IOC: it is entirely independent of the investor's preferences. It therefore doesn't have to be the original point of tangency between the IOC and the investor's indifference curves. This was point F in Figure 3; F is a long way from the action in Figure 4.

What this means for the investor is that, by borrowing or lending money, she can possibly reach a higher indifference curve than is possible just with her own money. Sharpe draws the indifference curves for three hypothetical individuals on Figure 4. The individual shown in Figure 3 had indifference curves tangential to point F, and this is now individual C in Figure 4.[90]

Individual C can reach a higher level of utility by borrowing money at the riskless rate P, and investing her own plus this borrowed money into asset F, to reach the risk/return combination signified by C* in Figure 4. This takes her from the indifference curve C1 to the higher indifference curve C3.

All together now?

At this stage, Sharpe encountered a problem. As well as every investor having a different set of indifference curves between risk and return, each would also have a different opinion about the return and risk that would be associated with each possible investment. Thus investor C might think that investment F – say the Internet company Yahoo – was likely to yield a low return at high risk, while investor A might expect that Yahoo will give high returns with little variation.

In other words, each investor would perceive a different 'cloud' of investment opportunities. The edge of the cloud of investment opportunities, the IOC, would be different for every investor, in terms of both location and the investments in it.

Equally, lenders may charge a different rate of interest to every borrower, so that the location P would differ between individuals. They might also restrict credit to some (or all) investors, so that the length of the line between each investor's P would differ – rather than being infinitely long, as Sharpe assumed. It might not even be a line, but could well be a curve, with lenders charging a higher rate of interest as borrowers committed themselves to more and more debt.

Sharpe circumvented these problems by (a) assuming that all investors could borrow or lend as much as they liked at the same rate, and (b)

assuming that investors all agreed on the expected prospects for each and every investment. Sharpe realised that these were extreme assumptions, but he justified them by an appeal to the methodological authority of Milton Friedman. In Sharpe's words:

> In order to derive conditions for equilibrium in the capital market we invoke two assumptions. First, we assume a common pure rate of interest, with all investors able to borrow or lend funds on equal terms. Second, we assume homogeneity of investor expectations: investors are assumed to agree on the prospects of various investments – the expected values, standard deviations and correlation coefficients described in Part II. Needless to say, these are highly restrictive and undoubtedly unrealistic assumptions. However, since the proper test of a theory is not the realism of its assumptions but the acceptability of its implications, and since these assumptions imply equilibrium conditions which form a major part of classical financial doctrine, it is far from clear that this formulation should be rejected – especially in view of the dearth of alternative models leading to similar results. (Sharpe 1964 [1991]; emphasis added)

Though Sharpe doesn't explicitly say this, he also assumes that investor expectations are accurate: that the returns investors expect firms to achieve will actually happen.

With these handy assumptions under his belt, the problem was greatly simplified. The point P was the same for all investors. The IOC was the same for all investors. Therefore all investors would want to invest in some combination of the riskless asset P and the same tangential investment f.

All that differed were investor risk/return preferences. Some would borrow money to move further 'northeast' (towards a higher return with higher risk) than the point at which their indifference map was tangential to the IOC. Others would lend money to move 'southwest' from the point of tangency between their indifference map and the IOC, thus getting a lower return and a lower risk.

All borrowing and lending movement would occur along the line Pf, and all investors would initially attempt to put their money into the investment f.

Since all investors will attempt to buy investment f, and no investors will attempt to buy any other investment, the market mechanism kicks in. Investment f rises in price, while all other investments fall in price. This process of re-pricing investments alters their returns, and flattens the edge of the IOC until such time as a large range of portfolios are tangential to P, as shown in Figure 5. These range from investment A, which has the lowest return and lowest risk, to B, which has the highest return and the highest risk. The process continues until such time as all investments exist in at least one portfolio lying along the new tangent line PZ.

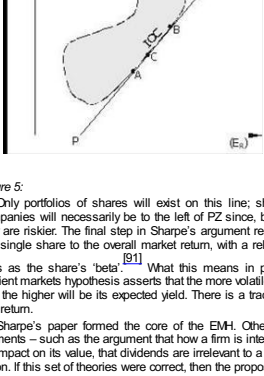


Figure 5:

Only portfolios of shares will exist on this line; shares of individual companies will necessarily be to the left of PZ since, being undiversified, they are riskier. The final step in Sharpe's argument relates the return on any single share to the overall market return, with a relation known these days as the share's 'beta'.[91] What this means in practice is that the efficient markets hypothesis asserts that the more volatile a share's returns are, the higher will be its expected yield. There is a trade-off between risk and return.

Sharpe's paper formed the core of the EMH. Others added ancillary elements – such as the argument that how a firm is internally financed has no impact on its value, that dividends are irrelevant to a share's value, and so on. If this set of theories were correct, then the propositions cited earlier would be true: the collective expectations of investors will be an accurate prediction of the future prospects of companies; share prices will fully reflect all information pertinent to the future prospects of traded companies;[92] changes in share prices will be entirely due to changes in information relevant to future prospects; and prices will 'follow a random walk', so that

past movements in prices give no information about what future movements will be.

Reservations

The outline above covers the theory as it is usually presented to undergraduates (and victims of MBA programs), and as it is believed by its adherents amongst stock brokers and speculators. But Sharpe was aware that it was unsatisfactory, mainly because of side-effects from the assumptions that investors are in complete agreement about the future prospects of traded companies, and that all investors can borrow or lend as much as they want at the riskless rate of interest.

One obvious side-effect of the first assumption is that, once equilibrium is reached, trade on the stock exchange should cease. Thereafter, any trading should merely be the result of the random arrival of new information, or the temporary disturbance of equilibrium via the floating of some new security. The trading profile of the stock market should therefore be like that of an almost extinct volcano.

Instead, even back in the 1960s when this paper was written, the stock market behaved like a very active volcano. It has become even more so since, and in 1987 it did a reasonable, though short-lived, impression of Krakatau. In 2000, we saw 25% movements in a week. October 2000 lived up to the justified reputation of that month during bull markets; heaven only knows how severe the volatility will be when the bubble finally bursts.

The second assumption implies that anyone could borrow sufficient money to purchase all the shares in, say, Microsoft, and pay no more than the riskless rate of interest to do it. This implies a degree of liquidity that is simply impossible in the real world.

Sharpe very honestly discussed both the reality of these assumptions, and the implications of dropping them. He readily conceded that "even the most casual empiricism" suggests that the assumption of complete agreement is false:

> People often hold passionately to beliefs that are far from universal. The seller of a share of IBM stock may be convinced that it is worth considerably less than the sales price. The buyer may be convinced that it is worth considerably more. (Sharpe 1970)

If this assumption is dropped, then in place of the single 'security market line' (the straight line in Figure 5), and a spectrum of efficient investments which is the same for all investors, there is a different security market line for each investor. The clean simplicity of the EMH collapses.

The assumption that we can all borrow (or lend) as much as we like at the riskless rate of interest is just as unrealistic as the assumption that all investors agree. Sharpe concedes that the theory collapses once one accepts the reality that the borrowing rate normally exceeds the lending rate, that investors are credit rationed, and that the borrowing rate tends to rise as the amount being borrowed increases:

> The consequence of accommodating such aspects of reality are likely to be disastrous in terms of the usefulness of the resulting theory . . . The capital market line no longer exists. Instead, there is a capital market curve – linear over some ranges, perhaps, but becoming flatter as [risk] increases over other ranges. Moreover, there is no single optimal combination of risky securities; the preferred combination depends upon the investors' preferences . . . The demise of the capital market line is followed immediately by that of the security market line. The theory is in a shambles. (Sharpe 1970; emphasis added)

But in the end, faced with a choice between an unrealistic theory and no theory at all, Sharpe opts for theory. His comfort in this choice continues to be Milton Friedman's methodological escape route that the unrealism of assumptions "is not important in itself. More relevant, the implications are not wildly inconsistent with observed behaviour" (Sharpe 1970).

But as discussed in Chapter 7, this argument that assumptions don't matter is only valid if they are negligibility assumptions (which dismiss features of the real world which are irrelevant or immaterial to the system being modelled) or heuristic assumptions (which are used to simplify argument en route to a more general theory, where the assumptions are dropped).

Do Sharpe's assumptions qualify under either of those headings? Clearly not. They are not negligibility assumptions – if they were, then dropping them would not leave the theory "in a shambles". They are not heuristic assumptions since, as Sharpe concedes, once they are dropped the theory collapses, and he had no alternative to offer.

Instead, they are domain assumptions (factors that are required to make the theory valid, and in the absence of which the theory is invalid), and therefore the theory is only valid in a world in which those assumptions apply.

That is clearly not our world. The EMH cannot apply in a world in which investors differ in their expectations, in which the future is uncertain, and in which borrowing is rationed. It should only have been taken seriously had

Sharpe or its other developers succeeded in using it as a stepping stone to a theory which took account of uncertainty, diverse expectations, and credit rationing. Since they did not do so, the EMH should never have been given any credibility – yet instead it became an article of faith for academics in finance, and a common belief in the commercial world of finance.

Sharpe deserves commendation for honestly discussing the impact on his theory of relaxing his assumptions – unfortunately, the same can't be said for the textbook writers who promulgated his views. However, the problems he saw with his theory are just the tip of the iceberg. There are so many others that it is difficult to think of a theory that could less accurately describe how stock markets behave.

Efficient or prophetic market?

Figure 6 illustrates the process which the EMH alleges investors use to determine the value of capital assets. Investors objectively consider information about the investment opportunities offered by different companies, and data about world economic prospects. Information that affects the future prospects of investments arrives randomly, generating random movements in the expected future prospects of firms. Investors' rational appraisal of this information leads to an efficient valuation of shares on the basis of expected return and risk, with price variations being caused by the random arrival of new information pertinent to share prices.

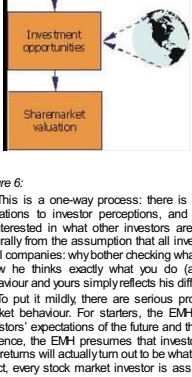


Figure 6:

This is a one-way process: there is no feedback from share market valuations to investor perceptions, and most importantly, investors are uninterested in what other investors are doing. This, of course, follows naturally from the assumption that all investors agree about the valuations of all companies: why bother checking what your neighbour thinks when you know he thinks exactly what you do (and any difference between his behaviour and yours simply reflects his different risk/return preferences)?

To put it mildly, there are serious problems with this theory of stock market behaviour. For starters, the EMH makes no distinction between investors' expectations of the future and the future which actually occurs. In essence, the EMH presumes that investors' expectations will be fulfilled: that returns will actually turn out to be what investors expected them to be. In effect, every stock market investor is assumed to be Nostradamus. What economists describe as 'efficient' actually requires that investors be prophetic.

As soon as you allow that investors can disagree, then this economic notion of 'efficient expectations' also collapses. If investors disagree about the future prospects of companies, then inevitably the future is not going to turn out as most – or perhaps even any – investors expect.

This divergence between expectations and outcomes will set up disequilibrium dynamics in the stock market – precisely the sort of behaviour that the EMH cannot model, because it is above all a theory of market equilibrium. If investors influence each other's expectations, this is likely to lead to periods when the market is dominated by pessimistic and optimistic sentiment, and there will be cycles in the market as its shifts from one dominant sentiment to the other.

The efficient markets hypothesis was used to berate market participants for believing that such phenomena as 'bull' and 'bear' markets actually existed: there was always only the efficient market. But even a slight concession to reality indicates that bull and bear phases will be part and parcel of a real-world stock market.

Risks ain't risks

Sharpe's measure of risk was standard deviation, a statistical measure of how much the values thrown up by some process vary. If values are fairly evenly distributed around an average, then roughly two-thirds of all

outcomes will be one standard deviation either side of the average.

For example, tests of IQ often have an average of 100 and a standard deviation of 16. This means that two thirds of the population will score between 84 and 116 on an IQ test.

There are at least two problems with applying this concept to investment:

Ÿ Is variability really what an investor means by risk?

Ÿ To actually work out a standard deviation, you need some process that has thrown up lots of historical data with a pattern which can be expected to continue recurring in the future.

Consider two investments: Steady has an average return of 3%, and a standard deviation of 3%; Shaky has an average return of 9%, and a standard deviation of 6%. Which one is 'riskier'?

According to Sharpe's criterion, Shaky is riskier: its standard deviation is twice as big as Steady's. However, according to any sane investor, Steady would be riskier – since there's a much higher chance of getting a negative return from Steady than there is from Shaky. In other words, what an investor really worries about is not so much variability as downside risk. Standard deviation is a very poor proxy for this, even if a standard deviation can be meaningfully calculated in the first place.

This brings us to the second problem. Standard deviation can be used as a measure of variability for things such as the expected outcome of a dice roll, the age at which someone dies, even a golfer's possible scores. However, even here there are differences in how reliable a guide historical averages and standard deviations can be to future outcomes. So long as a dice is well designed, a roll is going to have a one in six chance of turning up a 2 for a considerable time – until, for example, repeated rolls erode its edges. The historical averages for death, however, have changed dramatically in the West even during one lifetime, and major changes (for the better or worse, depending on whether genetic engineering or global ecological problems come out on top during the 21st century) can be expected in the future. And if an 18-year-old golfer had an average of 70 and a standard deviation of 5 now, would you rely on those numbers as a guide to his performance in 30 years' time?

In other words, for measures like standard deviation to be reliable, past outcomes must remain a reliable guide to future outcomes. This is not going to be the case for an investment, because the future performance of a company depends upon future economic circumstances, future inventions, the actions of future competitors, all things to which the past provides no reliable guide beyond a very short time horizon. Investment, and stock market speculation, are, in other words, subject not to risk, but to uncertainty.

We have already discussed the implications of uncertainty for economic analysis in chapters 7 and 9. For stock market investors, uncertainty means that the expected yield of an investment over the medium to long-term future simply can't be known:

> Our knowledge of the factors which will govern the yield of an investment some years hence is usually very slight and often negligible. If we speak frankly, we have to admit that our basis of knowledge for estimating the yield ten years hence of a railway, a copper mine, a textile factory, the goodwill of a patent medicine, an Atlantic liner, a building in the City of London, amounts to little and sometimes to nothing; or even five years hence. In fact, those who seriously attempt to make any such estimate are often so much in the minority that their behaviour does not govern the market. (Keynes 1936)

Uncertainty, not risk, is the main factor standing between investors and an accurate knowledge of the future prospects of companies. As a result, the expected yield of an investment, the other variable in the EMH model of investor behaviour, simply can't be known.

"The dark forces of time and ignorance…"

The efficient markets hypothesis argues that investors try to maximise their utility, where the only determinants of that utility are expected returns on the one hand, and risk on the other.

This kind of analysis has been soundly applied to interpret gambling. A gambler playing a game of blackjack faces known payoffs, and the known probabilities of drawing any given card. A good gambler is someone who intelligently applies these well-known regularities to decide how much to bet, when to hold, and when to risk another flip of the card.

This is an invalid concept to apply to an investor's behaviour, since the game played in the casino of the stock market is subject to uncertainty, not risk.

Nonetheless, investors still need to form some expectations of the future if they are going to act at all. These will be based partly on factors they currently know – such as prevailing economic conditions – and partly on factors they can't know. In practice, they rely mainly upon the knowable factors simply because they are knowable: investors therefore extrapolate current trends into the indefinite future. As Keynes puts it:

> It would be foolish, in forming our expectations, to attach great

weight to matters which are very uncertain. It is reasonable, therefore, to be guided to a considerable degree by the facts about which we feel somewhat confident, even though they may be less decisively relevant to the issue than other facts about which our knowledge is vague and scanty. For this reason the facts of the existing situation enter, in a sense disproportionately, into the formation of our long-term expectations; our usual practice being to take the existing situation and to project it into the future, modified only to the extent that we have more or less definite reasons for expecting a change. (Keynes 1936)

This is clearly an unreliable practice, but in an uncertain world there is simply no other way to act. It is something that we must do in order not to be paralysed into inaction, but it is something that, at a deep level, we are aware is untrustworthy. As a result, our forecasts of the future are tempered by an additional factor, the degree of confidence we have that these forecasts will be at least approximately correct. The more significant the degree of change expected, the more fragile that confidence will be.

The share market's valuations therefore reflect both collective forecasts, and the confidence with which these forecasts are made. In tranquil times these valuations will be relatively stable, but

> [i]n abnormal times in particular, when the hypothesis of an indefinite continuance of the existing state of affairs is less plausible than usual even though there are no express grounds to anticipate a definite change, the market will be subject to waves of optimistic and pessimistic sentiment, which are unreasoning and yet in a sense legitimate where no solid basis exists for a reasonable calculation. (Keynes 1936)

Therefore, in this uncertain world, the stock market will be ruled not by dispassionate analysis, but by euphoria, fear, uncertainty and doubt. It will be a place, not of analytic rationality, but of emotion.

The madness of the third degree

Keynes once described himself as a speculator who had lost two fortunes and made three. His assessment of the behaviour of stock market speculators was thus that of a well-informed insider. Keynes described the stock market as a game of

> Musical Chairs . . . a pastime in which he is victor . . . who secures a chair for himself when the music stops. These games can be played with zest and enjoyment, though all the players know that . . . when the music stops some of the players will find themselves unseated. (Keynes 1936)

The essence of this game is not to work out what particular shares are likely to be worth, but to work out what the majority of other players are likely to think the market will think they are worth, since "it is not sensible to pay 25 for an investment of which you believe the prospective yield to justify a value of 30, if you also believe that the market will value it at 20 three months hence" (Keynes 1936). In one of the most evocative analogies ever used by an economist, Keynes compared investing in shares to

> those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. (Keynes 1936)

Though this may seem to be a description of the behaviour of amateur investors in Internet chat rooms, Keynes insists that it is also the modus operandi of professional stock managers. Firstly, because the future is uncertain, the kind of long-term forecasting which the EMH assumes is the norm is effectively impossible. It is far easier to anticipate "changes in the conventional basis of valuation a short time ahead of the general public" (Keynes 1936).

Secondly, the boards that employ such professional stock managers discipline their behaviour to make them conform to the norm. Any manager who is truly trying to anticipate future economic trends is bound to make recommendations that are wildly at variance with what is popular in the market, and this behaviour will appear eccentric and ill-informed in comparison to the current market favourites. Imagine, for example, what would have happened to a funds manager who in mid-2000 advised the fund to sell all its shares in Yahoo, or Amazon, and spend the proceeds buying, for example, bonds.

As Keynes eloquently put it, "Worldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally." Unconventional managers are thus weeded out, leaving behind only those who swim with the crowd.

Thirdly, the long-term investor has to ignore the prospect of quick short-term capital gains, and this runs counter to human nature's desire for quick results.

Finally, a long-term investor can't afford to be highly geared, since the results of being wrong will be expensive and the accumulated financing cost over the long run will be great. Speculators, on the other hand, are attracted to gearing by the allure of large immediate gains now, at a cost of only minor short-term interest charges (especially when the prevailing aura of confidence during a bull market leads them to discount the possibility of large losses).

Thus, according to Keynes, rather than looking dispassionately at investment prospects and world economic conditions, the main thing share market investors do is look furtively and emotionally at each other, to attempt to predict how the majority will value particular companies in the immediate future.

This behaviour is pictured in Figure 7. Though investors do still keep an eye on individual investments and world conditions, and the world does throw in surprising events from time to time, in the main investors analyse the investment community itself.

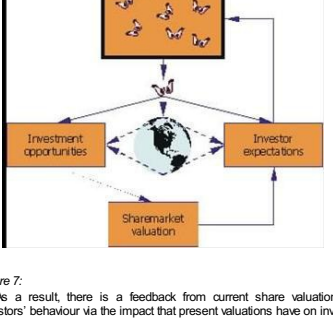


Figure 7:

As a result, there is a feedback from current share valuations to investors' behaviour via the impact that present valuations have on investor expectations. Arising market will tend to encourage investors to believe that the market will continue rising; a falling market will maintain the sentiment of the bears. Such a market can find itself a long way from equilibrium as self-reinforcing waves of sentiment sweep through investors. These waves can just as easily break – though long after any rational calculation might suggest that they should – when it becomes clear that the wave has carried valuations far past a level which is sustainable by corporate earnings.

So wrong it's almost right

The critiques above raise one curious question: how could a theory which was so obviously wrong nonetheless generate predictions about stock market behaviour that, at a superficial level, looked roughly right?

One of the key predictions of the EMH is that 'you can't beat the market': in a perfect capital market, price fluctuations simply reflect the random arrival of new information, and yesterday's price trends are as relevant to tomorrow's as the last roll of a dice is to the next.

On the other hand, if the market is as 'imperfect' as argued above, and trends therefore exist, surely it should be possible for the intelligent investor to profit from these trends? If so, wouldn't this eventually lead to all opportunities for profit being sought out, thus removing the trends and, hey presto, making the market efficient? Not necessarily, for two reasons: a factor discussed briefly in Chapter 8: 'chaos', and the institutional structure of the market, which Keynes detailed in the General Theory.

Alternatives

Different ways to 'think economically'

11. Finance and economic breakdown

Why stock markets crash

The efficient markets hypothesis says that the stock market's volatility is due to the random arrival of new information that affects the equilibrium value of shares. Allegedly, if it were not for the arrival of new information from outside the market, the market itself would be quiescent.

However, there is an alternative explanation that attributes most (though not all) of the market's volatility to its own internal dynamics.

Remarkably, these two explanations can predict statistical outcomes for share market prices that are almost indistinguishable from each other.

The kernel

If financial markets aren't efficient, then what are they? According to the 'fractal markets hypothesis', they are highly unstable dynamic systems that generate stock prices which appear random, but behind which lie deterministic patterns. According to the 'inefficient markets hypothesis', they are systems which overreact to good news and bad, leading to excessive asset price volatility which inhibits the performance of the real economy. According to the 'financial instability hypothesis', they are speculative vehicles existing in an uncertain environment, whose reactions to uncertainty over time generate financial cycles which drive the real economy into speculative booms and debt-driven busts. All three theories support the argument that unless finance markets are institutionally tamed, capitalism will remain subject to potentially catastrophic breakdown.

The roadmap

In this chapter I outline three different but consistent non-equilibrium theories of finance – the fractal markets hypothesis, inefficient markets hypothesis, and the financial instability hypothesis. The chapter concludes with a proposal that open-ended shares be replaced by time-limited 'voting bonds', in order to institutionally limit the degree of stock price instability.

The inherent instability of stock markets

The efficient markets hypothesis explains the price fluctuations that characterise financial markets as rational reactions by the markets to the random arrival of new information affecting the future prospects of companies. The three different approaches to finance outlined in this chapter all argue that these price fluctuations are due to the markets' own internal dynamics. These are two fundamentally different explanations for the same phenomenon: one based on exogenous shocks – the random arrival of external economic news – the other on internal dynamics – today's market prices being a reaction to yesterday's. How can two such different explanations account for the same data?

An analogy might help here. Some animal populations – for example, lemmings – are known to fluctuate wildly from year to year. There could be two explanations: the environment in which lemmings live could be so volatile that it causes extreme variations in population from one year to the next, and without this environmental volatility, lemming numbers could be constant. Or, the environment could be relatively stable, but the population dynamics of lemmings could be so volatile that it causes huge fluctuations in numbers from year to year.

Figure 1 shows a feasible yearly population pattern generated by a model of the first explanation. This model normalises the equilibrium lemmings population to 0.5, and has a standard deviation of the variation caused by the environment of 0.35.

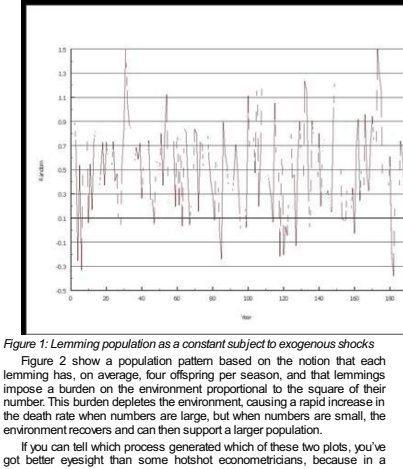


Figure 1: Lemming population as a constant subject to exogenous shocks

Figure 2 show a population pattern based on the notion that each lemming has, on average, four offspring per season, and that lemmings impose a burden on the environment proportional to the square of their number. This burden depletes the environment, causing a rapid increase in the death rate when numbers are large, but when numbers are small, the environment recovers and can then support a larger population.

If you can tell which process generated which of these two plots, you've got better eyesight than some hotshot econometricians, because in a comparison of statistical programs, one concluded that the numbers in Figure 2 were generated by the model like that in Figure 1 (Barnett 1999).[93]

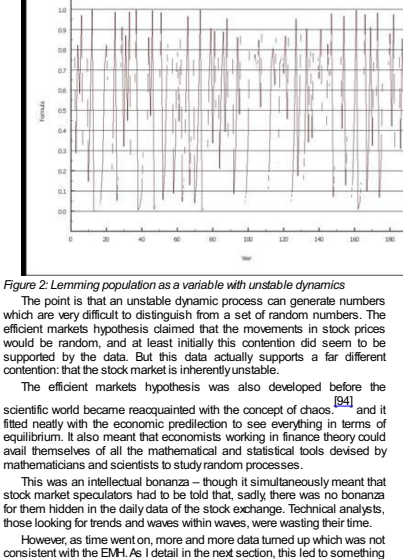


Figure 2: Lemming population as a variable with unstable dynamics

The point is that an unstable dynamic process can generate numbers which are very difficult to distinguish from a set of random numbers. The efficient markets hypothesis claimed that the movements in stock prices would be random, and at least initially this contention did seem to be supported by the data. But this data actually supports a far different contention: that the stock market is inherently unstable.

The efficient markets hypothesis was also developed before the scientific world became reacquainted with the concept of chaos,[94] and it fitted neatly with the economic predilection to see everything in terms of equilibrium. It also meant that economists working in finance theory could avail themselves of all the mathematical and statistical tools devised by mathematicians and scientists to study random processes.

This was an intellectual bonanza – though it simultaneously meant that stock market speculators had to be told that, sadly, there was no bonanza for them hidden in the daily data of the stock exchange. Technical analysts, those looking for trends and waves within waves, were wasting their time.

However, as time went on, more and more data turned up which was not consistent with the EMH. As I detail in the next section, this led to something

of a 'siege mentality' by supporters of the EMH, as they fought to defend their theory from attack. But it also inspired other researchers to develop alternative theories of stock market movements.

The fractal markets hypothesis

The fractal markets hypothesis is primarily a statistical interpretation of stock market prices, rather than a model of how the stock market, or investors in it, actually behave. Its main point is that stock market prices do not follow the random walk predicted by the EMH,[95] but conform to a much more complex pattern called a fractal. As a result, the statistical tools used by the EMH, which were designed to model random processes, will give systematically misleading predictions about stock market prices.

The archetypal set of random numbers is known as the 'normal' distribution, and its mathematical properties are very well known. A normal distribution with a average value of zero and a standard deviation of one will throw up a number greater than 1 15% of the time, a number greater than 2 just over 2% of the time, and a number greater than 3 only once every 750 times, and so on. The chance of a 'far from average' event occurring diminishes rapidly and smoothly the further the event is from the average.

The standard deviation of daily movements on the Dow Jones Industrial Average is roughly 1%. If stock market prices were generated by a normal process, then extreme movements – say a fall of more than 5% in just one day – would be vanishingly rare. The odds of any such event having occurred even once during the 20th century would be just over 1 in a 100.

> *In fact, there were over 60 such daily downward movements (and over 50 daily upward movements of 5% or more) during the 20th century.*

The fact that extreme movements occurred roughly 10,000 times more often than for a random process is fairly strong evidence that the process is not random at all (and there's lots more evidence besides this morsel).

A fractal set of numbers, on the other hand, is a far more pernicious beast. Specifically it is much more likely to generate extreme events than a normal distribution, and one large movement is likely to be followed by another large movement – another feature of stock markets which the EMH finds very difficult to explain.[96] A fractal pattern also displays 'self-similarity': the data pattern looks the same regardless of whether you are looking at a short data period – such as one day, or a week – or longer periods, such as a year or even a century.

The basic idea behind a fractal is that each number in the series is a simple but nonlinear function of previous numbers in the series. This was illustrated in the lemmings example: this year's number was exactly equal to 4 times last year's number, minus 4 times the square of last year's number. As you can see from Figure 2, this generates a wildly varying set of numbers, which could appear to be randomly generated. Yet in a true 'random number generator' such as a dice, the next number is independent of all previous numbers – rolling a 6 now doesn't change the odds of rolling a 6 on your next throw, they will still be 1 in 6.

Applying this to the stock market, it is quite possible that each price movement is a complex function of previous price movements.

This might seem to imply that, if the fractal markets hypothesis is correct, it should be easy to make money out of the stock market – in which case the hypothesis would be invalid, since it isn't easy to profit as a trader. However, there is another key aspect of fractal systems which comes into play here, which is known as 'sensitive dependence on initial conditions'.

Even if you knew precisely the 'system' which generated the Dow Jones Industrial Average, you could never know the precise value of the index because of rounding error. Let's say your initial measure of its value was out by 1/10th of a per cent – rather than being, say 10396.5, it was actually 10396.6.

One day (or iteration) later, your model would be wrong by (say) 1%; one day later by 10%; and a day after that, it would be completely useless as a means of predicting the following day's value. This is because any measurement errors you make in specifying the initial conditions of a fractal model grow exponentially with time, whereas for a random model the errors normally grow linearly (and can even fall with time for a stable system). As Ott puts this dilemma:

> The exponential sensitivity of chaotic solutions means that, as time goes on, small errors in the solution can grow very rapidly (i.e., exponentially). Hence, after some time, effects such as noise and computer roundoff can totally change the solution from what it would be in the absence of these effects. (Ott 1993)

Ott gives the example of a chaotic function called the Henon Map being simulated on a computer which is accurate to 15 decimal places: the smallest difference it can record between two numbers is 0.00000000000001. He shows that if your initial measurement of the system was out by precisely this much, then after 45 iterations of the model, your estimate of where the system is would be completely wrong. Attempting to overcome this problem by more computing power is futile:

> Suppose that we wish to predict to a longer time, say, twice as long. Then we must improve our accuracy by a tremendous amount, namely 14 orders of magnitude! In any practical situation, this is likely to be impossible. Thus, the relatively modest goal of an improvement of prediction time by a factor of two is not feasible. (Ott 1993)

Applying this to the stock market, it is possible to hold two apparently contradictory attitudes simultaneously: the market is driven largely by endogenous processes in which previous price movements determine future price movements; and it is impossible – or very difficult – to predict which way the market will move, and by how much.

Much of the fractal markets hypothesis is directed at critiquing the notion that price movements in the stock market are random – as I noted earlier, it is primarily a way to characterise the properties of the statistics the market throws up, rather than a theory of how the market actually behaves. However, it makes one important behavioural observation that runs directly counter to the EMH's assumptions about investors.

This is that the market will be stable when it allows investors with different time horizons to trade smoothly. As a result, heterogeneity – the fact that all investors are not the same – is a vital part of this theory. As Peters puts it:

> Take a typical day trader who has an investment horizon of five minutes and is currently long in the market. The average five-minute price change in 1992 was -0.000284 per cent [it was a 'bear' market], with a standard deviation of 0.05976 per cent. If, for technical reasons, a six standard deviation drop occurred for a five minute horizon, or 0.359 per cent, our day trader could be wiped out if the fall continued. However, an institutional investor – a pension fund, for example – with a weekly trading horizon, would probably consider that drop a buying opportunity because weekly returns over the past ten years have averaged 0.22 per cent with a standard deviation of 2.37 per cent. In addition, the technical drop has not changed the outlook of the weekly trader, who looks at either longer technical or fundamental information. Thus the day trader's six-sigma [standard deviation] event is a 0.15-sigma event to the weekly trader, or no big deal. The weekly trader steps in, buys, and creates liquidity. This liquidity in turn stabilizes the market. (Peters 1994)

The fractal markets hypothesis thus explains the stability of the market by the realistic assumption that traders differ in their time horizons. It also alleges that instability is likely to occur if all investors suddenly switch to the same time horizon.

The fractal markets hypothesis is thus more consistent with stock market data, more robust, and completely untainted by any assumption that the market is in, or tends towards, equilibrium. But it still doesn't provide an answer to what is actually generating the data: what is the system behind the fractal? To answer that question, we have to return to the kind of institutional analysis that Keynes provided in 1936. Two such analyses have been provided: by Robert Haugen in the 'inefficient markets hypothesis', and Hyman Minsky in the 'financial instability hypothesis'.

The inefficient markets hypothesis

After a long career as an academic finance economist, Bob Haugen presents the diametrically opposite case from the efficient markets hypothesis with gusto in three short books: The Beast on Wall Street, The New Finance, and The Inefficient Stock Market. Anyone who is or is thinking of speculating in the market – or is suffering from having done so – should read all three. Amid an extensive catalogue of data that contradicts the efficient markets hypothesis, Haugen presents the alternative case for "a noisy stock market that overreacts to past records of success and failure on the part of business firms, and prices with great imprecision" (Haugen 1999b).[97]

Though Haugen makes no reference to Keynes, the reasons he gives for the market behaving in this way echo the arguments Keynes made back in 1936, that in the real world of uncertainty, few if any stock market speculators trade on the basis of new information. Instead, they trade on the basis of how they think other market participants will, on average, expect the market to react to news. Unlike the efficient market hypothesis, this 'news' can include the most recent movements of stock prices themselves.

In fact, in today's stock market, the major news will always be the most recent movements in stock prices, rather than 'real' news from the economy.

Haugen argues that there are three sources of volatility: event-driven, error-driven, and price-driven (the efficient markets hypothesis models only the first, on the belief that the other two can't exist in the equilibrium of an efficient market). The second results from the market over-reacting to news, then over-adjusting itself once the initial mistake has become obvious.

The third is the phenomenon of the market reacting to its own volatility, building price movements upon price movements, in the same way that neighbourhood dogs can sometimes keep yelping almost indefinitely after one of them has started. Haugen argues that this endogenous instability

accounts for over three-quarters of all volatility.

He also argues that the market's endogenous instability has a severe and deleterious impact on the functioning of a modern capitalist economy.

Firstly, if the stock market has any role at all in directing investment funds, then its valuations will direct them very badly. Price-driven volatility will lead to some companies which will in the long term turn out to be worthless being given massive funding – which will then be wasted – while potentially worthy ventures will be starved of funds. According to Haugen, the managers of a firm which has been seriously over-valued by the market "overinvestConsumers get what they don't want". On the other hand, an under-valued firm "would invest to produce a product that consumers really want, if it could raise capital at a fair price, but in this market, it can't" (Haugen 1999a). Overall, by providing too much money to ventures which, in the long run, are going to turn out not to be all that profitable, while providing too little money to those which, in the long run, will be worthwhile, the market causes the economy to grow less rapidly and less smoothly than it could.

Secondly, as well as causing investment to be badly apportioned, Haugen argues that the stock market's endogenous volatility reduces the overall level of investment. Over the long term, the risk-free real rate of return has averaged about 1%, whereas the risk premium for investing in stocks has averaged about 6%. This means that investors have required a return of about 7% on their investments – with the result that investments which predict a lower expected rate of return don't get funded.

Haugen argues that investors require this higher return to compensate themselves for the risk of investing on the market, yet most of this risk results from the endogenous instability of the market itself. Since his statistical research indicates that price-driven volatility accounts for almost 95% of all volatility, he argues that this risk premium would be substantially lower – perhaps as low as just 0.4%, versus the 6% it has been historically (Haugen 1999a). If the risk premium could be reduced to this level, then the rate of investment would be substantially higher:

> Price-driven volatility has greatly inhibited investment spending over the years. Ultimately, it has acted, and acts, as a serious drag on economic growth. (Haugen 1999a)

At the individual level, Haugen argues that the market's tendencies to over-react to news, and to be consumed with endogenous instability in prices, provide opportunities for non-institutional investors to profit from the market. However, at the macroeconomic level, Haugen believes, as did Keynes, that the economy would benefit if the market were restrained. His recommendation, again very similar to Keynes's, is to reduce the length of the trading day, or to limit trading to just one computer-assisted auction per day. He hopes that this would eliminate the phenomenon of price volatility driven by the market reacting to its own every move.

The financial instability hypothesis

Haugen's argument is remarkable for the extent to which it reproduces Keynes's analysis in the 1930s, without ever referring to Keynes. Minsky, on the other hand, consciously saw himself as building upon Keynes's (and Fisher's) insights to develop a theory of how finance interacts with the economy to produce booms and busts, and occasionally, major depressions. His analysis is distinctly macroeconomic, since he focuses upon the systemic behaviour of the overall economy, rather than just the behaviour of the stock market.

Minsky's analysis commences at a time when the economy is doing well (the rate of economic growth equals or exceeds that needed to reduce unemployment), but firms are conservative in their portfolio management (debt to equity ratios are low and profit to interest cover is high), and this conservatism is shared by banks, who are only willing to fund cash-flow shortfalls or low-risk investments.

The cause of this high and universally practised risk aversion is the memory of a not-too-distant financial failure, when many investment projects foundered, many firms could not finance their borrowings, and many banks had to write off bad debts. Because of this recent experience, both sides of the borrowing relationship prefer extremely conservative estimates of prospective cash flows: their risk premiums are very high.

However, the combination of a growing economy and conservatively financed investment means that most projects succeed. Two things gradually become evident to managers and bankers: "Existing debts are easily validated and units that were heavily in debt prospered: it pays to lever" (Minsky 1977 [1982]). As a result, both managers and bankers come to regard the previously accepted risk premium as excessive. Investment projects are evaluated using less conservative estimates of prospective cash flows, so that with these rising expectations go rising investment and asset prices. The general decline in risk aversion thus sets off both growth in investment and exponential growth in the price level of assets, which is the foundation of both the boom and its eventual collapse.

More external finance is needed to fund the increased level of investment and the speculative purchase of assets, and these external funds are forthcoming because the banking sector shares the increased optimism of investors (Minsky 1970 [1982]). This results in an increase in the accepted

debt to equity ratio and falling profit to interest cover.[98] a decrease in liquidity, and accelerated growth of credit.

In a regulated environment this will manifest itself via the growth of non-bank financial intermediaries; in a deregulated environment, there will simply be a rapid increase in the money supply. Depending upon how the money supply is measured, these changes will be recorded as either an increase in the relevant measure of the money supply, or in an increase in the velocity of money circulation.

This marks the beginning of what Minsky calls "the euphoric economy" (Minsky 1970 [1982]), where both lenders and borrowers believe that the future is assured, and therefore that most investments will succeed. Asset prices are revalued upward as previous valuations are perceived to be based on mistakenly conservative grounds. Highly liquid, low-yielding financial instruments are devalued, leading to a rise in the interest rates offered by them as their purveyors fight to retain market share.

Financial institutions now accept liability structures both for themselves and their customers "that, in a more sober expectational climate, they would have rejected" (Minsky 1970 [1982]). The liquidity of firms is simultaneously reduced by the rise in debt to equity ratios, making firms more susceptible to increased interest rates. The general decrease in liquidity and the rise in interest paid on highly liquid instruments triggers a market-based increase in the interest rate, even without any attempt by monetary authorities to control the boom. However, the increased cost of credit does little to temper the boom, since the rate of asset price inflation means that anticipated yields from speculative investments far exceed prevailing interest rates.

The condition of euphoria also permits the development of an important actor in Minsky's drama, the 'Ponzi financier' (named after a 1920s embezzler who peddled both financial and real estate schemes).[99] These capitalists profit by trading assets on a rising market, and incur significant debt in the process. The servicing costs for Ponzi debtors exceed the cash flows of the businesses they own, but the capital appreciation they anticipate far exceeds the interest bill. They therefore play an important role in pushing up the market interest rate, and an equally important role in increasing the fragility of the system to a reversal in the growth of asset values.

Rising interest rates and increasing debt to equity ratios eventually affect the viability of many business activities, reducing the interest rate cover, turning projects which were originally conservatively funded into speculative ones, and making ones which were speculative Ponzi. Such businesses will find themselves having to sell assets to finance their debt servicing – and this entry of new sellers into the market for assets pricks the exponential growth of asset prices.

With the price boom checked, Ponzi financiers now find themselves with assets that can no longer be traded at a profit, and levels of debt that cannot be serviced from the cash flows of the businesses they now control. Banks that financed these assets purchases now find that their leading customers can no longer pay their debts – and this realisation leads initially to a further bank-driven increase in interest rates. Liquidity is suddenly much more highly prized, holders of illiquid assets attempt to sell them in return for liquidity. The asset market becomes flooded and the euphoria becomes a panic, the boom becomes a slump.

As the boom collapses, the fundamental problem facing the economy is one of excessive divergence between the debts incurred to purchase assets, and the cash flows generated by them – with those cash flows depending upon both the level of investment and the rate of inflation. The level of investment has collapsed in the aftermath of the boom, leaving only two forces that can bring asset prices and cash flows back into harmony: asset price deflation, or current price inflation. This dilemma is the foundation of Minsky's iconoclastic perception of the role of inflation, and his explanation for the stagflation of the 1970s and early 1980s.

Minsky argues that if the rate of inflation is high at the time of the crisis, then though the collapse of the boom causes investment to slump and economic growth to falter, rising cash flows rapidly enable the repayment of debt incurred during the boom. The economy can thus emerge from the crisis with diminished growth and high inflation, but few bankruptcies and a sustained decrease in liquidity. Thus though this course involves the twin 'bads' of inflation and initially low growth, it is a self-correcting mechanism in that a prolonged slump is avoided. However, the conditions are soon re-established for the cycle to repeat itself, and the avoidance of a true calamity is likely to lead to a secular decrease in liquidity preference.

If the rate of inflation is low at the time of the crisis, then cash flows will remain inadequate relative to the debt structures in place. Firms whose interest bills exceed their cash flows will be forced to undertake extreme measures: they will have to sell assets, attempt to increase their cash flows at the expense of their competitors, or go bankrupt. In contrast to the inflationary course, all three classes of action tend to further depress the current price level, thus at least partially exacerbating the original imbalance.

If assets are sold as going concerns, then those who buy them face a lower cost of capital, and can undercut their rivals in the current goods

market. If assets are broken up, then the commodities that constitute them compete at reduced prices with new output from competitors. If firms attempt to increase cash flows by reducing their markups, they will be matched by other competitors in similar circumstances, thus leading if anything to further reduced cash flows and a general reduction in the price level. If firms go bankrupt, their stocks and assets will be sold into depressed markets, thus further reducing current prices. The asset price deflation route is therefore not self-correcting but rather self-reinforcing, and is Minsky's explanation of a depression.

Thus while Minsky still sees inflation as having deleterious effects during a relatively stable period of economic growth, he perceives it in quite a different light during a time of crisis. The fundamental problem during a financial crisis is the imbalance between the debts incurred to purchase assets, and the cash flows those assets generate. A high rate of inflation during a crisis enables debts that were based on unrealistic expectations to be nonetheless validated, albeit over a longer period than planned and with far less real gain to the investors. Alow rate of inflation will mean that those debts cannot be met, with consequent domino effects even for investments that were not unrealistic.

The above sketch basically describes Minsky's perception of an economy in the absence of a government sector. With what he calls 'Big Government', the picture changes in two ways, due to the impact of fiscal deficits and of Central Bank interventions on corporate cash flows. With a developed social security system, the collapse in cash flows that occurs when a boom becomes a panic will be at least partly ameliorated by a fall in tax revenues and a rise in government spending – the classic automatic stabilisers, though this time seen in a more monetary light. The collapse in credit can also be tempered or even reversed by rapid action by the Central Bank to increase liquidity.

With both these forces operating in all Western economies since WWII, Minsky expected the conventional cycle to be marked by "chronic and accelerating inflation" (Minsky 1980 [1982]). However, by the end of the millennium, the cost pressures that coincided with the slump of the early 1970s had long since been eliminated, by decades of high unemployment and the diminution of OPEC's cartel power. Inflation fell even further in the aftermath to the 1987 stock market crash and the 1989–93 recession.

If a crisis does occur after the Internet Bubble finally bursts, then it could occur in a milieu of low inflation (unless oil price pressures lead to an inflationary spiral). Firms are likely to react to this crisis by dropping their margins in an attempt to move stock, or to hang on to market share at the expense of their competitors. This behaviour could well turn low inflation into deflation.

The possibility therefore exists that America could once again be afflicted with a debt deflation – though its severity could be attenuated by the inevitable increase in government spending that such a crisis would trigger. America could well join Japan on the list of the global economy's 'walking wounded' – mired in a debt-induced recession, with static or falling prices and a seemingly intractable burden of private debt.

Real reform?

Deregulation of the financial sector was not the sole cause of the financial instability of the past twenty years. But it has certainly contributed to its severity, by removing some of the limited constraints to cyclical behaviour which exist in a regulated system.

These deregulations were mooted as 'reforms' by their proponents, but they were in reality retrograde steps, which have set our financial system up for a real crisis. I can only hope that, if the crisis is serious enough, then genuine reform to the finance sector will be contemplated. Reform, of course, cannot make capitalism stable; but it can remove the elements of our corporate system which contribute most strongly to instability.

The major institutional culprit has to be the finance sector itself, and in particular the elements of the stock market which lead to it behaving more like a casino than a place of reasoned calculation.

Keynes and Haugen suggest restricting access to the stock market, and reducing its volatility by such deliberate measures as allowing only one 'Walrasian auction' every day. However, these reforms have the potential to excessively stifle liquidity – as Keynes noted – and there is no guarantee that such a reform would be sustained as memories of the crisis that inspired it receded.

A potentially more effective reform suggestion has been made by the Norwegian systems engineer Trond Andresen.[100] This is that the very nature of the stock certificate should be altered, from being a perpetuity as it is now to a time-limited 'voting bond'.

One reason why speculative excess can overwhelm the market is that the value of a stock is essentially open-ended. As America's stock market bubble has shown, any valuation is sustainable, so long as the expectation develops that the stock can be on-sold for a profit.

Bonds, on the other hand, have strict limits to how much they can rise or fall in value. In part these limits are imposed by the interest-payment

obligations written into a bond. But in part they also arise from the fact that, on its maturity, a bond is redeemed – the issuer of the bond repays the buyer the face value of the bond.

In effect, a bond is like a skipping rope – with its initial and its terminal value fixed. It can move down and up in value over its lifetime as prevailing interest rates rise above and below its coupon rate, and as speculators speculate about the future course of the economy, but ultimately the redemption date imposes a finite limit to any possible over-valuation.

Stock certificates could be redesigned to resemble bonds in this respect. A coupon rate would not be promised, and unlike a bond, a share certificate would confer some ownership and voting rights upon the buyer. But a maximum term could be set for the certificate, at which time the certificate would be redeemed by the issuing company at its face value.

This system would hopefully reduce the volatility that dominates the secondary market in shares, and put the emphasis upon the primary market function, of raising capital for firms with new ventures. As a fundamental change in the nature of the ownership of corporations, it is more likely to stick than restrictions on trading hours, and the like. The last 20 years have shown how such sensible controls can come to be regarded as unwarranted restrictions on the market during a period of economic growth. A sensible reform should not only address the immediate problem, but should also cement itself into the fabric of society. Andresen's proposal may well have those attributes.

It is also no more radical than many other reforms that have been forced upon capitalism's chief casino after previous financial collapses. All the legal requirements that are now placed upon publicly listed companies – the requirement for prospectuses, the existence of a Securities Commission, etc. – were developed in response to excesses uncovered during previos speculative frenzies. Surely, when the Internet Bubble really bursts, it will be time to admit that one fundamental excess of the market as currently organised is its ability to allow sky-high valuations to develop.

So what?

What can I say? By promulgating the efficient markets hypothesis, which is predicated on each investor having the foresight of Nostradamus, economic theory has encouraged the world to play a dangerous game of stock market speculation. When that game comes unstuck, America in particular will most likely find itself as badly hobbled by debt as Japan has been for the past decade. This speculative flame may have ignited anyway, but there is little doubt that economists have played the role of petrol throwers rather than firemen. When crisis strikes, conventional economists will be the last people on the planet who can be expected to provide sage advice on how to return to prosperity – unless, as often happens in such circumstances, they drop their theoretical dogmas in favour of common sense.

For whom the bubble bursts

When the Great Crash of 1929 led to the Great Depression of the 1930s, many of the erstwhile heroes of the finance sector found themselves in the docks. It is unlikely that any particular economists will find themselves so arraigned, but there is little doubt that economic theory has been complicit in encouraging America's investing public to once again delude itself into a crisis.

Only time will tell whether the bursting of the Internet Bubble will lead to as dire an outcome as the Great Depression. Certainly, on many indicators, the 1990s bubble has left its septuagenarian relative in the shade. The price to earnings ratio peaked at over one and a half times the level set in 1929, the private and corporate debt to output ratio is possibly three times what it was prior to the Great Crash, and prices, though rising in some sectors, are generally quiescent. On all these fronts, Fisher's debt-deflation theory of great depressions seems a feasible outcome.

On the other hand, Minsky argued that 'Big Government' could stabilise an unstable economy, by providing firms with cash flow from which their debt commitments could be financed despite a collapse in private spending. Certainly, the US government of 2000 is 'big' when compared to its 1920s counterpart, and its automatic and policy interventions will probably attenuate any economic crash to something far milder than the Great Depression. What appears more likely for post-Internet America is a drawn-out recession like that experienced by Japan since its Bubble Economy collapsed in 1990.

12. Don't shoot me, I'm only the piano

Why mathematics is not the problem

Many critics of economics have laid the blame for its manifest failures at the feet of mathematics. Mathematics, they claim, has led to an excessive formalism in economics, which has obscured the inherently social nature of the subject.

While it is undeniable that an inordinate love of mathematical formalism has contributed to some of the intellectual excesses in economics, generally this reaction is as erroneous as blaming the piano for the discordant notes of a bad piano player. If anything should be shot, it is the pianist, not the piano.

Though mathematics has definite limitations, properly used, it is a logical tool that should illuminate, rather than obscure. Economists have obscured reality using mathematics because they have practised mathematics badly, and because they have not realised the limits of mathematics.

The kernel

If you divided the world's population into those who dislike mathematics, those who like it, and those who were indifferent, I suspect that 95% would be in the first camp, 5% in the second, and 0% in the third. Neoclassical economists come almost exclusively from the 'like it' camp, and therefore their arguments are almost always expressed in mathematical form. Most critics of economics come from the 'dislike it' camp, and frequently criticism of mathematics per se forms part of their criticism of economics.

Call me weird: I'm a critic of neoclassical economics who likes mathematics. But I am not alone. There are numerous mathematically inclined critics of conventional economics, and in many ways this book was written to convey their critiques to a non-mathematical audience. Not only is it possible to simultaneously like mathematics and dislike mainstream economics, but a sound knowledge of mathematics makes you an even more confirmed opponent – because much of the mathematics in conventional economic theory is unsound.

At one level, it is unsound because conditions which economists assume contradict other conditions needed for their models, so that the theory is built on a mathematical error. For example, the economic argument in favour of small competitive firms over monopolies assumes that the supply curve for a monopoly is identical to the aggregate supply curves of numerous competitive firms, and that both supply curves slope upwards because of rising marginal cost. As discussed in the extension to Chapter 4 on the web at Size/More, these two assumptions are mathematically inconsistent.

At a more profound level, conventional economics is mathematically unsound because it has not learnt the lesson which true mathematicians have learnt in the last century: that there are limits to mathematical logic.

The roadmap

In this chapter I argue that conventional economics has abused mathematics in two main ways: by practising bad mathematics, and by not acknowledging the limitations of mathematics. Many economic theorems result in logical contradictions which economists fail to recognise as such, and many other theorems are derived by falsely assuming that different quantities are in fact equal. Modern mathematics has also realised that there are limits to mathematical logic, but economists have evaded this realisation by effectively but unintentionally isolating themselves from mainstream mathematical science.

Bad mathematics

In a classic instance of 'those who live by the sword die by the sword', the school of economics that prides itself on being mathematical is subject to the indictment that its mathematics is erroneous. There are numerous theorems in economics that rely upon mathematically fallacious propositions. There are basically four ways in which this manifests itself in economic theory:

- Logical contradiction, where the theory is allegedly 'saved' by an assumption which in fact contradicts what the theory purports to show;
- Omitted variables, where an essential aspect of reality must be ignored to derive the mathematical results that economists wish to prove;
- False equalities, where two things that are not quite equal are treated as if they are equal; and
- Unexplored conditions, where some relation is presumed without exploring what conditions are needed to make this relation feasible.

Logical contradiction

The case outlined in Chapter 2 – the failed attempt to aggregate individual preferences to form community preferences with the same properties – is an excellent example of logical contradiction.

The economic theory of consumer behaviour begins with the proposition that it is possible to aggregate individual utility to derive a community preference map that has the same characteristics as the individual map. Economists have proven that this is only possible when the so-called Sonnenshein-Mantel-Debreu conditions apply: (a) that all consumers have the same preference map; and (b) that preferences do not change with income.

Condition (a) effectively means that there is only one consumer. Condition (b) effectively means that there is only one commodity. Aggregation is therefore strictly possible if there is only one consumer and only one commodity.

Clearly this is not aggregation at all.

A good mathematician would recognise this as proof by contradiction (Franklin & Daoud 1988). This is a clever technique where, to prove a statement, you assume its opposite and then show a contradiction. Therefore the statement is true.

For example, consider the problem that confronted Pythagoras and friends when they tried to work out the length of the hypotenuse of a right-angled triangle whose other sides were both 1 unit long. According to Pythagoras's theorem that 'the square of the length of the hypotenuse is equal to the sum of the squares of the other sides', this meant that the length of the hypotenuse was the square root of 2.

These Ancient Greeks initially believed that all numbers were 'rational', which meant that any number could be expressed as the fraction of two integers: thus 3Ú2, for example, is 3 divided by 2. But they could never accurately measure the value of the square root of 2 in terms of the ratio of two integers: every more accurate measurement led to a different fraction.

The reason that they couldn't find the 'right' two integers is that the square root of 2 is an irrational number: it can't be defined as the ratio of two integers.

This can be proven quite easily using proof by contradiction. You start with the opposite assumption – that it is possible to express the square root of 2 as a ratio of two integers. You then work on from this point, to show this leads to a contradiction. Therefore you show that the square root of 2 is irrational.[101]

The proofs which led to the Sonnenshein-Mantel-Debreu conditions can easily be described in the same fashion. You wish to prove that it is not possible to aggregate individual preferences to derive community preferences which display the same characteristics as individual preferences.

You start with the opposite assumption – that it is possible to aggregate the preferences of 2 or more different consumers over 2 or more different commodities to yield a community preference map. You then show that this is only possible if there is only 1 individual and only 1 commodity. This contradicts your starting assumption that there are multiple different consumers and different commodities. Therefore you have proven, by contradiction, that it is not possible to aggregate individual preferences to derive community preferences.

The trouble is that economists were hoping that they could prove that it was possible to aggregate. In this sense, they were in the same situation as the ancient Pythagoreans, who were trying to prove that all numbers were rational – they were not at all pleased to find that they were wrong.

At least the Pythagoreans relented: they abandoned the belief that all numbers were rational, and accepted that there were numbers which could not be described as the ratio of two integers. Mathematics thus absorbed the existence of irrational numbers, and went on from there to many other discoveries.

Economists, on the other hand, have been unwilling to abandon their concept of rationality. Faced with an equivalent discovery – that society cannot be understood as the sum of the rational individuals within it – economics has instead enshrined these and similar logical contradictions as "intuitively reasonable" (Gorman 1953) abstractions that are needed to forge a link between individual and collective utility.

This is bad mathematics. It has led to bad economics, which has avoided the more complex but richer visions of the economy that flow from coming to terms with these disproofs of simplistic notions.

Omitted variables

Bad mathematics also shows up in such hallowed economic concepts as maximising profit by equating marginal cost and marginal revenue. As is shown in Chapter 3, this mantra of the everyday economist is only true if we ignore time. Once time is rightly included in the analysis, this mantra loses its validity: it is then mathematically evident that, to maximise profits over time, a firm should ensure that its marginal revenue exceeds its marginal cost.

Many critics of conventional economics have previously argued that time is the most crucial variable left out of economic analysis, but most of these critics have then eschewed mathematics itself as a result. Few of them have realised that conventional economics is not just poor economics, but also poor mathematics. Good mathematical economics incorporates time as an essential aspect of reality, and results in a type of economic analysis that is profoundly different from conventional neoclassical economics.

Time is not the only variable omitted by economists, of course. Other notable examples include uncertainty, and the formation of expectations under uncertainty, from the analysis of macroeconomics, as discussed in Chapter 10.

False equalities

One popular but erroneous step in conventional economic argument is to assert that something that is extremely small can be treated as zero. For example, this argument is used to 'derive' the horizontal demand curve facing the competitive firm. It is said that such a firm is so small that it has no influence on the market price, and therefore takes the market price as given. With each firm having zero impact on the market, and with each of them facing a horizontal demand curve, the competitive market as a whole is said to achieve the economic Nirvana of equating the marginal cost of producing its output to the marginal benefit society enjoys from consuming that output.

The web extension to Chapter 4 at More/Size shows that this argument is fallacious. If the competitive industry as a whole faces a smoothly downward-sloping demand curve, then every firm in it must also face a (however slightly) downward-sloping demand curve. It is not correct to treat an infinitesimally small amount as zero.

This is especially so when economists then pretend to 'aggregate up' from the individual firm to derive a result that applies at the aggregate level. What results is not mathematical analysis, but a mathematical sleight of hand – the intellectual equivalent of a magician's trick.

The model of perfect competition illustrates this nicely. The argument starts with the market having a downward-sloping demand curve and an upward-sloping supply curve. Step one of the trick is to omit showing the downward-sloping marginal revenue curve, which must be there if the market demand curve slopes downward, and which is distinctly different from, and steeper than, the demand curve. Step two is to break the market demand curve into lots of tiny bits, each of which must also slope downwards if the whole curve slopes downwards, but to persuade the audience that the slope of each of these little lines is so flat that they can be treated as horizontal. Hence for the individual firm, the demand curve and the marginal revenue curve are identical. The final stage of the trick is to return to the market level by adding up all the individual firm's marginal cost curves, and to show that price is set by the intersection of the demand curve and the supply curve. The troublesome market marginal revenue curve has been made to disappear, and the trick is complete.

The special irony of this piece of magic is that the magician doesn't realise that a trick is being pulled. Economists are so used to presuming that an infinitesimal amount is equivalent to zero that they don't even realise they are breaching one of the fundamental rules of mathematics.

Unexplored conditions

There are numerous examples of this phenomenon. The comparison of monopoly to perfect competition presumes that the supply curve for the competitive industry is equivalent to the marginal cost curve for the monopoly. Chapter 4 showed that this is only possible if the two 'curves' are the same horizontal straight line. The theory of the labour market presumes that the supply curve of labour is upward sloping; Chapter 5 showed that this is not a necessary outcome of the neoclassical theory of labour supply. The analysis of production requires that the money value of capital is an adequate measure of the amount of capital; Chapter 6 showed that it is not, once we acknowledge that machines are produced by other machines and labour.

That these (and doubtless many other) logical conundrums exist indicates that economists do not explore the logical foundations of their beliefs. This in itself is not necessarily unscientific, as discussed in Chapter 7, it is a standard practice that scientists in a given school within a science do not challenge what Lakatos describes as the 'hard core' of their beliefs. But it is a sign of how fragile the neoclassical hard core is that so many elements of it can be shown so easily to be internally inconsistent.

It is also unscientific that, when such logical flaws are either pointed out by critics (as with Sraffa's critique in Chapter 6) or discovered by believers (as with the Sonnenschein-Mantel-Debreu conditions in Chapter 2), neoclassical theory adopts 'ancillary assumptions' which are clearly absurd (such as the 'machines as putty' notions which were put forward during the debate with Sraffa and his supporters, or the SMD conditions used to save the theory of consumption, which amount to assuming that all consumers have identical, income-independent tastes). This, to Lakatos, is the sign of a degenerative scientific research program which is preoccupied with

adjusting its ancillary beliefs to defend its hard core, whereas a truly progressive program would be expanding the range of real-world phenomena which its theory explains. The school of economics which gives pride of place to the word 'rational' hardly displays rational behaviour when its core beliefs are challenged. I expect that the new logical conundrums pointed out in this book will generate further displays of irrational behaviour by conventional economists.[102]

Mathematics is therefore not the reason why conventional economics is so bad. Instead, bad economics is supported by bad mathematical practice. But this is only half of story about how economics has abused mathematics. Economics has also accidentally inoculated itself against many of the advances of modern mathematics. One essential aspect of modern mathematics that economics has not realised is that mathematics today has a humility that economics lacks, because mathematicians have proven that mathematics has limits.

The limits to mathematics

Economics remains perhaps the only area of applied mathematics that still believes in Laplace's dictum that, with an accurate enough model of the universe and accurate enough measurement today, the future course of the universe could be predicted.

For mathematicians, that dictum was dashed in 1899 by Poincar–'s proof of the existence of chaos. Poincar– showed that not only was it impossible to derive a formula which could predict the future course of a dynamic model with three or more elements to it, but even any numerical approximation to this system would rapidly lose accuracy. The future could only be predicted if the present was known to infinite accuracy, and this was clearly impossible.

Today, mathematicians are quite comfortable with the proposition that most mathematical problems cannot be explicitly solved in a manner which yields the kind of didactic statements which economics makes as a matter of course – such as 'perfect competition gives superior welfare outcomes to monopoly', 'free trade is superior to protection', and so on. Such definitive pronouncements are generally only possible when the problem is essentially the same as working out where two straight lines intersect. This class of models is known as algebraic.

Some algebraic equations are rather difficult to solve because there is no standard formula to apply. But there are standard formulas available to solve systems of algebraic equations where all the equations are 'straight lines'. This is the type of mathematics which economic theory generally tries to apply to economic problems.

However, this body of mathematics is an appropriate model of only a tiny subset of real-world systems – and that subset does not include true economic analysis.[103]

The more appropriate starting point for mathematical models of the economy are dynamic equations, in which the relationships between variables cannot be reduced to straight lines. These are known in mathematics as nonlinear differential equations. The vast majority of these cannot be solved, and once three or more such equations interact, they are impossible to solve.

Table 1 summarises the situation. Economic theory attempts to analyse the economy using techniques appropriate to the upper left-hand part of Table 1 (with boldface text), when in fact the appropriate methods are those in the lower right-hand part (with cells shaded gray).

	Linear		Non-linear		
Equations	One equation	Several equations	Many equations	One equation	Several equatio
Algebraic	Trivial	Easy	Possible	Very difficult	Very difficult
Ordinary Differential	Easy	Difficult	Essentially impossible	Very difficult	Impossi
Partial Differential	Difficult	Essentially impossible	Impossible		

Table 1: The solvability of mathematical models (adapted from Costanza 1993)

Other developments, such as Godel's proof that a mathematical system cannot be self-contained – so that it must take some axioms on faith – and the proof that there were some mathematical problems which could not be solved, added to this realisation by mathematicians and physicists that mathematics and science had innate limits. As a result, in place of Laplace's grand conceit, there is a humility to modern mathematics. The future cannot be known, mathematics cannot solve every problem, some things may not be knowable.

But these epiphanies passed economists by: they continue to believe in a clockwork universe, in which a proper specification of the conditions of today could enable you to predict the future for all time. Nowhere is this

vanity more obvious than in the school's defining works, Walras' Elements of Pure economics and Debreu's Theory of Value.

Walras' arrogance towards those economists who would not practise mathematics is still the typical attitude today held by economists towards those who criticise their use of mathematics:

> As for those economists who do not know any mathematics, who do not even know what is meant by mathematics and yet have taken the stand that mathematics cannot possibly serve to elucidate economic principles, let them go their way repeating that "human liberty will never allow itself to be cast into equations" or that "mathematics ignores frictions which are everything in social science" and other equally forceful and flowery phrases. They can never prevent the theory of the determination of prices under free competition from becoming a mathematical theory. Hence, they will always have to face the alternative either of steering clear of this discipline and consequently elaborating a theory of applied economics without recourse to a theory of pure economics or of tackling the problems of pure economics without the necessary equipment, thus producing not only very bad pure economics but also very bad mathematics. (Walras 1900 [1954])

As this book has shown, it is economists who have been guilty of very bad mathematics. But just as important is the fact that they do not appreciate the limits to mathematics.

At least Walras could be forgiven for not being aware of Poincar–'s theorem of 1899 – though he had sought out Poincar– in a forlorn attempt to garner support for his mathematisation of economics. Debreu's opus predated the rediscovery of chaos by Lorenz, but that he could even conceive of modelling the economy as a system in which all production and exchange decisions were "made now for the whole future", and in which the theory of uncertainty was "free from any probability concept and formally identical with the theory of certainty", betrayed a profound lack of appreciation of the mathematics of his day (not to mention the real world).

The modern manifestation of this ignorance of the limits of mathematics is a widespread – though not universal – failure by economists to appreciate the importance of nonlinear analysis and chaos theory. If I had a cent for every time I heard an economist comment that 'chaos theory hasn't amounted to much' – well, I wouldn't be wealthy, but I could afford an expensive meal or two.

Chaos theory has 'not amounted to much' in economics because its central tenets are antithetical to the economic obsession with equilibrium. In other sciences, chaos theory, complexity analysis and their close cousin evolutionary theory have had profound impacts. It shows how isolated economics has become from the scientific mainstream of the late 20th and early 21st century that such ignorant views could be commonplace.

The recurring nightmare of straight lines

Virtually every critique detailed in this book has led to the result that some relationship between phenomena that economics argued was curved had to instead be a straight line.

The economic theory of consumer behaviour argued that a person's consumption of a commodity could change in any direction as his/her income rose: if it was a luxury, consumption would rise relative to other commodities; if a necessity, consumption could fall. Instead, the Sonnenshein- Mantel-Debreu conditions show that if the theory is to aggregate from the individual to the market demand curve, Engels curves must be straight lines.

The economic theory of production argues that output is subject to diminishing marginal returns, so that as the variable input rises, output rises less than proportionately – the relationship is curved. Sraffa's critique shows that, in general, output should rise proportionately: the relationship should be a straight line.

The economic theory of competition argues that perfect competition is superior to monopoly. But the only conditions under which the comparison is watertight involve a straight line relationship between inputs and outputs – not the curved relationship asserted by the concept of diminishing marginal productivity.

Why do straight lines haunt economic theory, when it is forced to be logical?

The answer to this dilemma has a lot to do with one of the basic notions of economics, the belief that society is no more than the sum of its parts. This asserts that to work out the whole, all you have to do is add the parts up. This requires that the interactions between the parts are either zero or negligible. The only interaction that one variable can have with another is the one economists want to use, simple addition: your utility plus my utility equals social utility; your output plus my output equals industry output; and so on.

This categorically rules out interactions where one variable is multiplied by another (where both are likely to be large numbers). One obvious such interaction occurs in working out a firm's revenue. A firm's revenue equals the number of units it sells, times the sale price. Economics argues that the

quantity a firm will sell is a function of price – to invoke a higher supply from the firm, the price has to rise to offset the effect of diminishing marginal productivity.

If both the price and the quantity are treated as variables, then the firm's revenue is equal to one variable (price) times another (quantity). This can't be allowed if economists are to treat society as no more than the sum of its parts, so economists assume that the price a competitive firm faces is a constant. Then the firm's revenue equals a constant (price) times a variable (quantity).

However, economists then want it both ways: they want price for the entire industry to be a variable, but price for the individual firm to be a constant, without the firms interacting in any way. This just can't happen mathematically – as discussed in Chapter 4. So if they force this situation by making an invalid assumption, it inevitably means that something else that they want to treat as a variable has to instead be treated as a constant. Hence the recurrent nightmare of a straight line.

The future of mathematics in economics

There is little doubt that the close identification of neoclassical economics with mathematical analysis has given mathematics a bad name amongst critical economists, and members of the ordinary public who are critical of economics.

This is likely to lead to a backlash against mathematics in economics, if the discipline ever rids itself of its dominance by neoclassical economics. This would be a great pity, since, as I hope this book has shown, properly used, mathematical reasoning debunks unsound economics. Furthermore, with its limitations fully appreciated, it and computer simulations can assist in the construction of sound alternative analyses. But if mathematics is avoided for its own sake, in reaction to how economics embraced it for its own sake, then the development of a meaningful economics will be stymied.

We now turn to some of the alternatives to conventional economics that do exist – warts and all. We begin with the most radical alternative – Marxian economics. You may, if you have typecast me as 'left-wing', expect me to praise Marxian analysis. If so, you are in for a surprise.

13. Nothing to lose but their minds

Why most Marxists are irrelevant, but most of Marx is not

Marxian economics is clearly one of the alternatives to the neoclassical way of 'thinking economically', and by rights I should be discussing it in the next chapter, which looks at alternatives to conventional economics. However, in an illustration of the fact that conservative economists do not have a monopoly on unsound analysis, Marxian economics is hobbled by a logical conundrum as significant as any of those afflicting neoclassical economics.

This conundrum has split non-orthodox economists into two broad camps. One tiny group continues to work within what they see as the Marxian tradition, and spends most of its time trying to solve this conundrum. The vast majority largely ignore Marx and Marxian economics, and instead develop the schools of thought discussed in the next chapter.

The kernel

One defining belief in conventional Marxian economics is that labour is the only source of profit: while machines are necessary for production, labour alone generates profit for the capitalist. This proposition is a key part of the radical appeal of Marxism, since it argues that capitalist profit is based upon exploitation of the worker.

Marxists argue that labour is the only source of profit because it is the only commodity where one can distinguish between 'commodity' and 'commodity-power'. When any other commodity was sold, the purchaser takes it lock, stock and barrel. But with labour, the capitalist 'purchaser' does not own the worker. Instead, he pays a subsistence wage, which can be represented by a bundle of commodities; this is the cost of production of the ability to work, which Marxists describe as the commodity 'labour-power'. The capitalist then puts the labourer to work for the length of the working day, during which time the worker produces a different bundle of commodities that is worth more than his subsistence wage. The difference between the output of labour and the cost of maintaining labour-power is the source of profit.

Since no such distinction can be made for machinery, the capitalist 'gets what he paid for' and no more when he buys a machine, whereas with labour, he gets more than he paid for. Therefore machines only transfer their value to the product.

This proposition has been shown to lead to severe logical problems, so the vast majority of critical economists have in practice abandoned Marx's logic. However, a minority of economists continue to swear allegiance to what they perceive as Marx's method, and continue to strive to invent ways in which the proposition that labour is the only source of profit can be maintained.

The critiques which have been made of this notion on mathematical grounds are cogent, but have been challenged by Marxian economists on philosophical or methodological grounds. However, there are philosophical reasons why the proposition that labour is the only source of profit are invalid, and these reasons were first discovered by Marx himself. Unfortunately, Marx failed to properly understand his own logic, and instead preserved a theory that he had in fact shown to be erroneous. Once Marx's logic is properly applied, his economics becomes a powerful means of analysing a market economy – though not one which argues that capitalism must necessarily give way to socialism. Unfortunately, given the ideological role of Marxism today, I expect that Marxian economists will continue to cling to an interpretation of Marx that argues for capitalism's ultimate demise.

The roadmap

In this chapter I explain the classical economics concept of 'value', and the manner in which Marx developed this into the labour theory of value. I illustrate the logical problems with the proposition that labour is the only source of value. I then outline Marx's brilliant philosophical analysis of the commodity, and show that this analysis contradicts the labour theory of value by arguing that all inputs to production are potential sources of value.

Marxian economics and the economics of Marx

If a 19th century capitalist Machiavelli had wished to cripple the socialist intelligentsia of the 20th century, he could have invented no more cogent weapon than the labour theory of value. Yet this theory was the invention, not of a defender of capitalism, but of its greatest critic: Karl Marx.

Marx used the labour theory of value to argue that capitalism harboured an internal contradiction, which would eventually lead to its downfall and replacement by socialism. However, Marx's theory itself had an internal contradiction that would invalidate Marx's critique of capitalism if it could not be resolved. Consequently, solving this enigma became the 'Holy Grail' for Marxist economists. Whereas 19th century revolutionaries spent their time

attempting to overthrow capitalism, 20th century revolutionaries spent theirs attempting to save the labour theory of value. Capitalism itself had no reason to fear them.

Despite valiant efforts, Marxist economists failed in their quest – and they achieved little else. As a result, while Marx's thought still has considerable influence upon philosophers, historians, sociologists and left-wing political activists, at the beginning of the 21st century, Marx and Marxists are largely ignored by other economists.[104] Most non-orthodox economists would acknowledge that Marx made major contributions to economic thought, but it seems that overall Samuelson was right: Marx was a 'minor Post-Ricardian' – someone who took classical economics slightly further than had David Ricardo, but who ultimately led it into a dead end.

This conclusion is false. Properly understood, Marx's theory of value liberates classical economics from its dependence on the labour theory of value, and makes it the basis for a deep and critical understanding of capitalism. But in a truly Machiavellian irony, the main factor obscuring this richer appreciation of Marx is the slavish devotion of Marxist economists to the labour theory of value.

To see why Marx's theory of value is not the labour theory of value, we have to first delve into the minds of the great classical economists Adam Smith and David Ricardo.

Value – a prelude

The proposition that something is the source of value raises two questions: what is 'value' anyway, and why should any one thing be the source of it?

A generic definition of value – one which encompasses the several schools of thought in economics which have used the term – is that value is the innate worth of a commodity, which determines the normal ('equilibrium') ratio at which two commodities exchange. One essential corollary of this concept is that value is unrelated to the subjective valuation which purchasers put upon a product. In what follows, I'll use 'value' in this specific sense, not in any of its more colloquial senses.

The classical economists also used the terms 'value in use' (or 'use-value') and 'value in exchange' (or 'exchange-value') to distinguish between two fundamental aspects of a commodity: its usefulness, and the effort involved in producing it. Value in use was an essential aspect of a commodity – why buy something which is useless? – but to the classical economists, it played no role in determining price.

Their concept of usefulness was also objective, focussing upon the commodity's actual function rather than how it affected the user's feelings of well-being. The use-value of a chair was not how comfortable it made you feel, but that you could sit in it.

In contrast, the neoclassical school argues that value, like beauty, is 'in the eye of the beholder' – that utility is subjective, and that the price, even in equilibrium, has to reflect the subjective value put upon the product by both the buyer and the seller. Neoclassical economics argues that the equilibrium ratio at which two products exchange is determined by the ratio of their marginal utilities to their marginal costs.

As we have seen in Chapters 2 and 3, there are serious problems with the economic theory of pricing. But it has some appeal in comparison to the classical approach, since it seems reasonable to say that price should be determined both by the innate worth of a product, however that is defined, and by the buyer's subjective valuation of it.

The general classical reply to this concept was that, sure, in the short run and out of equilibrium, that would be true. But the classical school was more interested in 'long run' prices, and in the prices of things which could easily be reproduced.

In the long run, price would be determined by the value of the product, and not by the subjective valuations of the buyer or seller. For this reason, the classical school tended to distinguish between price and value, and to use the former when they were talking about day to day sales, which could be at prices which were above or below long run values.

As well as having some influence out of equilibrium, subjective utility was the only factor that could determine the value of rare objects. As Ricardo put it:

> There are some commodities, the value of which is determined by their scarcity alone. No labour can increase the quantity of such goods, and therefore their value cannot be lowered by an increased supply. Some rare statues and pictures, scarce books and coins, wines of a peculiar quality, which can be made only from grapes grown on a particular soil, of which there is a very limited quantity, are all of this description. Their value is wholly independent of the quantity of labour originally necessary to produce them, and varies with the varying wealth and inclinations of those who are desirous to possess them. (Ricardo 1817)

Thus where scarcity was the rule, and the objects sold could not easily be reproduced, price was determined by the seller's and buyer's subjective

utilities. But this minority of products was ignored by the classical economists.

Marx gave an additional explanation of why in a developed capitalist economy, the subjective valuations of both buyer and seller would be irrelevant to the price at which commodities exchanged.

This was the historical argument that, way back in time, humans lived in small and relatively isolated communities, and exchange between them was initially a rare and isolated event. At this stage, the objects being exchanged would be items that one community could produce but the other could not. As a result, one community would have no idea how much effort had gone into making the product, and the only basis for deciding how to exchange one product for another was the subjective valuation that each party put upon the products. As Marx put it:

> The exchange of commodities, therefore, first begins on the boundaries of such communities, at their points of contact with other similar communities, or with members of the latter. So soon, however, as products once become commodities in the external relations of a community, they also, by reaction, become so in its internal intercourse. The proportions in which they are exchangeable are at first quite a matter of chance. What makes them exchangeable is the mutual desire of their owners to alienate them. Meantime the need for foreign objects of utility gradually establishes itself. The constant repetition of exchange makes it a normal social act. In the course of time, therefore, some portion at least of the products of labour must be produced with a special view to exchange. From that moment the distinction becomes firmly established between the utility of an object for the purposes of consumption, and its utility for the purposes of exchange. Its use-value becomes distinguished from its exchange-value. On the other hand, the quantitative proportion in which the articles are exchangeable, becomes dependent on their production itself. (Marx 1867; emphasis added)

The most famous example of two products being exchanged on the basis of the perceived utility rather than their underlying value is the alleged exchange of the island of Manhattan for a bunch of beads.[105] This price would never have been set if trade between the Dutch and the Indians had been a long established practice, or if the Indians knew how little work it took to produce the beads.

In an advanced capitalist nation, factories churn out mass quantities of products specifically for exchange – the seller has no interest in the products her factory produces. The sale price reflects the cost of production, and the subjective utility of the buyer and seller are irrelevant to the price.[106]

There is thus at least a prima facie plausibility to the argument that value alone determines the equilibrium ratio at which commodities are exchanged. The problem comes with the second question: what is the source of value?

Physiocrats

The first economists to systematically consider this question.[107] answered that the source of all value was land.

The argument, in a nutshell, was that land existed before man did. Therefore man – or more specifically, man's labour – could not be the source of value. Instead, value came from the land as it absorbed the energy falling on it from the sun. Man's labour simply took the naturally generated wealth of the land and changed it into a different form. Land generated a surplus, or net product, and this enabled both growth and discretionary spending to occur.

Manufacturing, on the other hand, was 'sterile': it simply took whatever value the land had given, and transformed it into different commodities of an equivalent value. No formal proof was given of this latter proposition, beyond an appeal to observation:

> Maxims of Economic Government. I: Industrial work does not increase wealth. Agricultural work compensates for the costs involved, pays for the manual labour employed in cultivation, provides gains for the husbandmen, and, in addition, produces the revenue of landed property. Those who buy industrial goods pay the costs, the manual labour, and the gain accruing to the merchants; but these goods do not produce any revenue over and above this. Thus all the expenses involved in making industrial goods are simply drawn from the revenue of landed property– no increase of wealth occurs in the production of industrial goods, since the value of these goods increases only by the cost of the subsistence which the workers consume. (Quesnay cited in Meek 1962; emphasis added)

Since land determined the value of commodities, and the price paid for something was normally equivalent to its value, the ratio between the prices of two commodities should be equivalent to the ratios of the land needed to produce them.

Smith (and Ricardo)

The physiocratic answer to the source of value reflected the school's origins in overwhelmingly rural France. Adam Smith, a son of Scotland and neighbour to the 'nation of shopkeepers', was strongly influenced by the physiocrats. But in The Wealth of Nations (which was published in the year in which the first steam engine was installed) Smith argued that labour was the source of value. In Smith's words:

> The annual labour of every nation is the fund which originally supplies it with all the necessaries and conveniences of life which it annually consumes, and which consist always either in the immediate produce of that labour, or in what is purchased with that produce from other nations. (Smith 1776)

The growth of wealth was due to the division of labour, which increased because the expansion of industry allowed each job to be divided into ever smaller specialised sub-tasks. This allowed what we would today call economies of scale: an increase in the size of the market allowed each firm to make work more specialised, thus lowering production costs (his most famous example of this was of a pin factory; this passage, which is better known than it is read, is reproduced on the web at Marx/Mbre).[108]

Smith therefore had an explanation for the enormous growth in output which occurred during the Industrial Revolution. However, he had a dilemma: for reasons discussed below, Smith knew that, though labour was the source of value, it could not possibly determine price. Yet value was supposed to determine the ratio at which two commodities exchanged.

The dilemma arose because two commodities could only exchange on the basis of the amount of direct labour involved in their manufacture if only labour was required for their production. Smith gave the example of exchange in a primitive hunting society:

> In that early and rude state of society which precedes both the accumulation of stock and the appropriation of land, the proportion between the quantities of labour necessary for acquiring different objects seems to be the only circumstance which can afford any rule for exchanging them for one another. If among a nation of hunters, for example, it usually costs twice the labour to kill a beaver which it does to kill a deer, one beaver should naturally exchange for or be worth two deer. (Smith 1776)

However, once there had been an 'accumulation of stock' – once a market economy had evolved – then paying for the labour alone was not sufficient: the price had also to cover profit:

> As soon as stock has accumulated in the hands of particular persons, some of them will naturally employ it in setting to work industrious people, whom they will supply with materials and subsistence, in order to make a profit by the sale of their work, or by what their labour adds to the value of the materials. In exchanging the complete manufacture either for money, for labour, or for other goods, over and above what may be sufficient to pay the price of the materials, and the wages of the workmen, something must be given for the profits of the undertaker of the work who hazards his stock in this adventure. (Smith 1776)

So Smith was forced to concede that the price had to be high enough to pay for not just the hours of labour involved in making something, but also a profit. For example, if the deer hunter was an employee of a deer hunting firm, then the price of the deer had to cover the hunter's labour, and also a profit margin for the firm.

The problem became more complicated still when land was involved. Now the price had to cover labour, profit, and rent. Smith's statement of this reveals that this 'father of economics' was rather more cynical and critical of market relations than some of his descendants:

> As soon as the land of any country has all become private property, the landlords, like all other men, love to reap where they never sowed, and demand a rent even for its natural produce– (Smith 1776)

In the end, Smith was reduced to an 'adding up' theory of prices: the price of a commodity represented in part payment for labour, in part payment for profit, and in part payment for rent. There was therefore no strict relationship between value and price.

Ricardo

Though he paid homage to his predecessor, Ricardo was, to say the least, critical of Smith's treatment of the relationship between value and price. He began his Principles of Political Economy and Taxation (Ricardo 1817) with an emphatic statement of the belief he shared with Smith, that labour was the determinant of the value of a commodity:

> The value of a commodity, or the quantity of any other commodity for which it will exchange, depends on the relative quantity of labour which is necessary for its production. (Ricardo 1817)

However, he was much more aware than Smith of the need for precise definitions, and of the difficulties in going from value to price.

Smith had used two measures of the amount of labour contained in a product: 'labour embodied' and 'labour commanded'. Labour embodied was the amount of direct labour time it actually took to make a commodity. Labour commanded, on the other hand, was the amount of labour-time you could buy using that commodity.

If, for example, it took one day for a labourer to make a chair, then the chair embodied one day's labour. However, that chair could well sell for an amount equivalent to two day's wages – with the difference accounted for by profit and rent. The chair would therefore command two days' labour.

Ricardo argued that the former measure was far less volatile than the latter. He believed, in common with most classical economists, that workers received a subsistence wage. Since this would always be equivalent to a fairly basic set of commodities – so much food, clothing, and rental accommodation – it would not change much from one year to the next. The latter measure, however, reflected the profit earned by selling the worker's output, and this would vary enormously over the trade cycle.

His solution for the value/price dilemma was that the price of a commodity included not just direct labour, but also the labour involved in producing any tools. Ricardo took up Smith's deer and beaver example and elaborated upon it. Even in Smith's example, some equipment had to be used to kill the game, and variations in the amount of time it took to make the equipment would affect the ratio in which deer and beavers were exchanged:

> Even in that early state to which Adam Smith refers, some capital, though possibly made and accumulated by the hunter himself, would be necessary to enable him to kill his game. Without some weapon, neither the beaver nor the deer could be destroyed, and therefore the value of these animals would be regulated, not solely by the time and labour necessary to their destruction, but also by the time and labour necessary for providing the hunter's capital, the weapon, by the aid of which their destruction was effected.
>
> Suppose the weapon necessary to kill the beaver, was constructed with much more labour than that necessary to kill the deer, on account of the greater difficulty of approaching near to the former animal, and the consequent necessity of its being more true to its mark; one beaver would naturally be of more value than two deer, and precisely for this reason, that more labour would, on the whole, be necessary to its destruction. (Ricardo 1817)

Thus the price of any commodity reflected the labour which had been involved in creating it, and the labour involved in creating any means of production used up in its manufacture. Ricardo gave many numerical examples in which the labour involved in producing the means of production simply reappeared in the product, whereas direct labour added additional value over and above its means of subsistence – because of the difference between labour embodied (which equalled a subsistence wage) and labour commanded (which included a profit for the capitalist).[109]

However, Smith and Ricardo were both vague and inconsistent on key aspects of the theory of value.

Though he generally argued that labour was the source of value, on several occasions Smith counted the work of farm animals as labour.[110] Though he failed to account for the role of machinery in the creation of value, he also argued that machines could produce more value than it took to produce them – which would mean that machinery (and animals) would be a source of value, in addition to labour:

> The expense which is properly laid out upon a fixed capital of any kind, is always repaid with great profit, and increases the annual produce by a much greater value than that of the support which such improvements require. (Smith 1776)

Ricardo more consistently implied that a machine added no more value to output than it lost in depreciation, but he also occasionally lapsed into completely ignoring the contribution of machinery to value.[111]

Marx's labour theory of value

Where his forebears implied and were vague, Marx stated and was emphatic: labour was the only source of value, in the sense that it could add "more value than it has itself" (Marx 1867). Marx called this difference between the value embodied in a worker and the value the worker added to production 'surplus value', and saw it as the sole source of profit.

He was critical of Ricardo for not providing an explanation of why this difference existed – in Ricardo's terms, for not having a systematic explanation of why labour embodied differed from labour commanded. As Marx put it:

> Ricardo starts out from the actual fact of capitalist production. The value of labour is smaller than the value of the product which it

creates–The excess of the value of the product over the value of the wages is the surplus-value– For him, it is a fact, that the value of the product is greater than the value of the wages. How this fact arises, remains unclear. The total working-day is greater than that part of the working day which is required for the production of wages. Why? That does not emerge. (Marx 1861 Part II)

The best that Ricardo could offer, Marx claimed, was that:

> [t]he value of labour is therefore determined by the means of subsistence which, in a given society, are traditionally necessary for the maintenance and reproduction of the labourers.
>
> But why? By what law is the value of labour determined in this way?
>
> Ricardo has in fact no answer, other than–the law of supply and demand–He determines value here, in one of the basic propositions of the whole system, by demand and supply – as Say notes with malicious pleasure. (Marx 1861 Part II)

Similarly, Marx rejected Smith's musings on the productivity of machinery, and concurred with Ricardo that a machine only added as much value to output as it lost through depreciation:

> The maximum loss of value that they can suffer in the process, is plainly limited by the amount of the original value with which they came into the process, or in other words, by the labour-time necessary for their production. Therefore, the means of production can never add more value to the product than they themselves possess independently of the process in which they assist. However useful a given kind of raw material, or a machine, or other means of production may be, though it may cost £150, or, say 500 days' labour, yet it cannot, under any circumstances, add to the value of the product more than £150. (Marx 1867)

Marx likewise concurred with Ricardo's definition of value, cited above, that it "depends on the relative quantity of labour which is necessary for its production". Value in turn determined the price at which commodities exchanged, with commodities of an equivalent value – commodities containing an equivalent quantity of labour[112] – exchanging for the same price (in equilibrium).

This exchange of equivalents nonetheless still had to enable capitalists to make a profit, and Marx was disparaging of any explanation of profits which was based on 'buying cheap and selling dear':

> To explain, therefore, the general nature of profits, you must start from the theorem that, on the average, commodities are sold at their real values, and that profits are derived by selling them at their values, that is, in proportion to the quantity of labour realised in them. If you cannot explain profit upon this supposition, you cannot explain it at all. (Marx 1847)

Marx gave two explanations for the origin of surplus value. One was a 'negative' proof, by a process of elimination based on the unique characteristics of labour. The other was a 'positive' proof, based on a general theory of commodities. Most Marxist economists are aware of only the negative proof.

The origin of surplus value (I)

This was that labour was a unique commodity, in that what was sold was not actually the worker herself (which would of course be slavery), but her capacity to work, which Marx called labour-power. The value (or cost of production) of labour-power was the means of subsistence, since that is what it took to reproduce labour-power. It might take, say, 6 hours of labour to produce the goods which are needed to keep a worker alive for 1 day.

However, what the capitalist actually received from the worker, in return for paying for her labour-power, was not the worker's capacity to work (labour-power), but actual work itself. If the working day was 12 hours long (as it was in Marx's day), then the worker worked for 12 hours – twice as long as it actually took to produce his value. The additional 6 hours of work was surplus labour, which accrued to the capitalist and was the basis of profit. As Marx put it:

> The labourer receives means of subsistence in exchange for his labour-power; the capitalist receives, in exchange for his means of subsistence, labour, the productive activity of the labourer, the creative force by which the worker not only replaces what he consumes, but also gives to the accumulated labour a greater value than it previously possessed. (Marx 1847)

This difference between labour and labour-power was unique to labour: there was no other commodity where 'commodity' and 'commodity-power' could be distinguished. Therefore other commodities used up in production simply transferred their value to the product, whereas labour was the source of additional value. Surplus value, when successfully converted into money by the sale of commodities produced by the worker, was in turn the source

of profit.

The labour theory of value and the demise of capitalism

This direct causal relationship between surplus value and profit meant there was also a direct causal relationship between what Marx called the rate of surplus-value and the rate of profit.

The rate of surplus value was the ratio of the surplus labour-time performed by a worker to the time needed to reproduce the value of labour-power. In our example above, this ratio is 1 to 1, or 100%: 6 hours of surplus labour to 6 hours of what Marx called necessary labour.

Marx defined the rate of profit as the ratio of surplus (which he denoted by the symbol s) to the sum of the inputs needed to generate the surplus. Two types of inputs were needed: necessary labour, and the means of production (depreciation of fixed capital plus raw materials, intermediate goods, etc.). Marx called necessary labour variable capital (for which he used the symbol v), because it could increase value, and he called the means of production used up constant capital (for which he used the symbol c), because it could not increase value.

Taking the example of weaving which Marx used extensively, during 1 working day a weaver might use 1000 yards of yarn and wear out 1 spindle. The yarn might have taken 12 hours of (direct and indirect) labour to make, and the spindle the same. Thus the sum of the direct labour-time of the worker, plus the labour-time embodied in the yarn and the spindle, is 36 hours: 12 hours' labour by the weaver, 12 for the yarn, and 12 for the spindle. The ratio of the surplus to c + u is 6:30 for a rate of profit of 20%.

Marx assumed that the rate of surplus value – the ratio of s to v – was constant, both across industries and across time.[113] Simultaneously, he argued that the competitive forces of capitalism would lead to capitalists replacing direct labour with machinery, so that for any given production process, c would get bigger with time. With s/v constant, this would decrease the ratio of s to the sum of c and v, thus reducing the rate of profit.

Capitalists would thus find that, regardless of their best efforts, the rate of profit was falling.[114] Capitalists would respond to this by trying to drive down the wage rate, which would lead to revolt by the politically aware working class, thus leading to a socialist revolution.[115]

Well, it was a nice theory. The problem was that, even if you accepted the premise that labour was the only source of value, the theory still had major logical problems. Chief among these was what became known as the transformation problem.

The transformation problem

The transformation problem arises from the fact that capitalists are motivated not by the rate of surplus value, but by the rate of profit. If the rate of surplus value is constant across industries, and labour is the only source of surplus, then industries with a higher than average ratio of labour to capital should have a higher rate of profit. Yet if a capitalist economy is competitive, this situation cannot apply in equilibrium, because higher rates of profit in labour-intensive industries should lead to firms moving out of capital-intensive industries into labour-intensive ones, in search of a higher rate of profit.

Marx was not an equilibrium theorist, but this problem was serious because his description of equilibrium was inconsistent. Somehow, he had to reconcile a constant rate of surplus value across industries with at least a tendency towards uniform rates of profit.

Marx's solution was to argue that capitalism was effectively a joint enterprise, so that capitalists earned a profit which was proportional to their investment, regardless of whether they invested in a labour-intensive or capital-intensive industry:

> Thus, although in selling their commodities the capitalists of the various spheres of production recover the value of the capital consumed in their production, they do not secure the surplus-value, and consequently the profit, created in their own sphere by the production of these commodities--So far as profits are concerned, the various capitalists are just so many stockholders in a stock company in which the shares of profit are uniformly divided per 100. (Marx 1894)

He provided a numerical example (Marx 1894) that purported to show that this was feasible. He first provided a table (Table 1) showing the production of surplus value by a number of industries with differing ratios of variable to constant capital (in modern terms, varying labour to capital ratios).

Capitals	Constant Capital	Variable Capital	Rate of Surplus Value	Surplus Value	Product c
I	80	20	100%	20	120
II	70	30	100%	30	130
III	60	40	100%	40	140

IV	85	15	100%	15	115
V	95	5	100%	5	105
Sum	390	110	100%	110	610

Table 2: Marx's unadjusted value creation table, with the rate of profit dependent upon the variable to constant ratio in each sector

In this 'value' table, a higher ratio of labour to capital is associated with a higher rate of profit. Thus 'labour-intensive' industry III, with a labour to capital ratio of 2:3, earns the highest 'value' rate of profit of 40%, while 'capital-intensive' industry V, with a 1:20 ratio, makes a 'value' rate of profit of just 5%.

Then Marx provided a second table in which the same industries earned a uniform rate of profit, now in terms of price rather than value. In contrast to Table 1, now all industries earned the same rate of profit.

The numbers in this example appeared feasible. The sums are consistent: the sum of all prices in Table 2 (610) equals the sum of the value created in Table 1; the sum of surplus value in Table 1 equals the sum of the differences between input costs (500) and the price of output in Table 2 (610). But this apparent consistency masks numerous internal inconsistencies. The best proof of this was provided by the Sraffian economist Ian Steedman (this next section is unavoidably technical, and can be skipped at first reading).

Capitals	Constant Capital	Variable Capital	Surplus Value	Price	Rate of Profit
I	80	20	22	122	22%
II	70	30	22	122	22%
III	60	40	22	122	22%
IV	85	15	22	122	22%
V	95	5	22	122	22%
Sum	390	110	110	610	22%

Table 3: Marx's profit distribution table, with the rate of profit now uniform across sectors

Marxist economics after Sraffa

We have already seen in Chapter 6 the damage Sraffa's crucible did to the economic theory of price determination and income distribution. In an illustration of the comparatively non-ideological nature of Sraffian analysis, Steedman showed that Sraffa's method could equally well critique Marxian economics.

The basis of Sraffa's system is the acknowledgment that commodities are produced using other commodities and labour. Unlike conventional economics – which has invented the fictional abstraction of 'factors of production' – Marx's system is consistent with Sraffa's 'production of commodities by means of commodities' analysis (indeed, Marx's economics was a major inspiration for Sraffa).

Steedman began with an illustrative numerical model of an economy with just three commodities: iron, corn and gold. Iron and labour were needed to produce all three commodities, but neither gold nor corn were needed to produce anything.[116] Table 3 shows the quantities of inputs and outputs in Steedman's hypothetical economy.

Industries	Inputs		Outputs		
	Iron	labour	Iron	Gold	Corn
Iron	28	56	56		
Gold	16	16		48	
Corn	12	8			8
Totals	56	80	56	48	8

Table 4: Steedman's hypothetical economy

The numbers in this table represent arbitrary units: the iron units could be tonnes, the labour units hours, gold units ounces, and corn units bushels – and any other set of arbitrary units would do as well. However, since each input is measured in a completely different unit, the numbers only add up down the columns: they don't add across the rows.

To analyse the labour theory of value, Steedman first had to convert these into the 'labour-value' units which Marx used. For simplicity, he set the labour-value of one unit ('hour') of labour at 1. Converted into value terms, Table 4 then says that it takes 28 times whatever the 'labour-value' of a tonne iron is, plus 56, to produce 56 times whatever the 'labour-value' of a tonne of of iron is. A bit of simple algebra shows that one tonne of iron has a labour-value of 2.

	c	v	s	Totals
Iron	56	14	42	112

Gold	32	4	12	48
Corn	24	2	6	32
Totals	112	20	60	192

Table 5: Steedman's physical table in Marx's value terms

Similar calculations show that the labour-value of an ounce of gold is 1, and the labour-value of a bushel of corn is 4.

The next stage in the analysis is to work out the value of the commodity labour-power. It might appear that this has already been done – didn't he set this equal to 1? No, because this represents the total amount of labour performed, and in Marx's theory, workers get paid less than this. They get paid, not for their contribution to output, but for the commodity labour-power, whose value is equal to the means of subsistence.

Steedman assumed that it took 5 bushels of corn to reproduce the labour used in this hypothetical economy. Therefore the total value of labour-power in the entire economy was equal to the labour-value of 5 bushels of wheat. Since a bushel of wheat has a labour-value of 4, this means that the value of labour-power across the whole economy was 20 (and therefore, one unit of labour had a labour-value of 1Ú4). The difference between this amount and the total labour performed – 80 hours of labour, which we have set equal 80 units of labour-value – is surplus value. So v, in Marx's scheme, is 20, while s is 60, for a rate of surplus value of 300%.

These numbers now allow the physical input data in Table 3 to be converted into Marx's labour-value terms. Since Marx assumed that the rate of surplus value was the same across all industries, 1Ú4 of the labour input in each industry represents v, while 3Ú4 represents s. Taking the iron industry, of the 56 labour-value units of direct labour, 14 represent v and 42 represent s. Since Table 5 is now in consistent units (everything is measured in labour-value units), the table adds up both horizontally and vertically.

	Inputs						
	c	v	Total	Profit Rate	Markup	Total Price	Per unit price
Iron price of production	56	14	70	45%	31.82	101.82	1.82
Gold price of production	32	4	36	45%	16.36	52.36	1.09
Corn price of production	24	2	26	45%	11.82	37.82	4.73
	112	20	132		60	192	

Table 6: Steedman's prices table in Marx's terms

With this table constructed, we can now calculate the average rate of profit in Marx's terms – which is the ratio of total s to the sum of c and v, or 60/132 (this factors to 5/11, and is equal to a rate of profit of 455Ú11%). In equilibrium, this rate of profit will apply across all industries, since otherwise capitalists would be shifting their resources from one sector to another. Steedman then multiplied the input values by 1 plus this uniform rate of profit to yield Marx's 'transformation' of values into prices.

So far, so good. Just as with Marx's table, the sum of values equals the sum of prices, and the sum of profits equals the sum of surplus values. However, all is not as well as it seems.

Table 5 tells us that the price of the total output of the iron industry is 101.82 (let's call this dollars, even though in these models the price simply means the ratios in which commodities exchange). If we divide this by the physical output of 56 tonnes of iron, then this means the price per tonne is $1.82. If the iron industry pays this price for its iron inputs in the next period, it will pay out $50.91. To hire the workers it needs, it has to buy sufficient corn: the amount works out to 3.5 bushels (this is the total amount of corn consumed by all workers [5 bushels] multiplied by the fraction of the total workforce employed in the iron industry). This costs $16.55. The iron industry's total outlays are thus $67.46, and yet (if Marx's equilibrium price calculations are accurate), it can sell its output for $101.82, for a profit of $34.36. But this is $2.55 more than the profit in the previous period.

Clearly there is an inconsistency – or rather, at least one. The simplest is that Marx converted the output into price terms, but didn't convert the inputs. However, it's worse than this: even if you amend this error, you get nonsense results: what is supposed to be an equilibrium (and therefore stationary) turns out not to be stationary at all.

Steedman then shows that you don't have to 'transform' physical quantities into values, and values into prices: you can instead derive prices directly from the physical data and the equilibrium assumption of a uniform rate of profit. The basis of this is that, in equilibrium, the prices have to enable each sector to just pay for its inputs and make the average rate of profit. Thus for the iron industry, the price of its 28 tonnes of iron inputs, plus the price of its 56 hours of labour, plus the standard markup, must just

equal the price of its 56 tonnes of iron output. There are two similar equations for corn and gold, and one final relation linking the wage to the cost of the subsistence amount of corn. If the gold price is notionally set to $1, this yields the average rate of profit, wage, and prices of iron and corn (in terms of gold) shown in Table 6.

Variable	Value; Price in terms of gold
Rate of profit	52%
Iron price of production	1.71
Gold price of production	1
Corn price of production	4.3

Table 7: Profit rate and prices calculated directly from output/wage data

Things don't look so good for Marx's tables now. Firstly, the rate of profit and prices worked out directly from the data (in Table 6) differ from those derived by taking Marx's route through the concept of value (in Table 5). Worse, whereas Marx's numbers aren't consistent – they are supposed to describe an equilibrium situation, but don't – the numbers derived directly from the data are consistent.

Take iron, for example. The iron sector pays $1.71 per tonne for its 28 tonnes of inputs, for a total of $47.88. It buys 3.5 bushels of wheat for $4.3 a bushel, for an outlay of $15.05. Total expenses of production are therefore $62.93. It then marks this up by the rate of profit to a total of $95.65. Except for the effect of rounding error, this equals the price of iron ($1.71) times the output (56 tonnes).[117]

Steedman concluded that, far from value determining prices, prices could not be accurately derived from values. Instead, prices could be worked out directly from the physical production data, and knowledge of the real wage: value calculations were both superfluous and misleading. He concluded that

> [t]here is no problem of transforming values into prices, etc., to be solved. The 'transformation problem' is a 'non-problem', a spurious problem which can only be thought to arise and to have significance when one is under the misapprehension that the rate of profit must be determined in terms of labour quantities. Once it is seen that there is no such necessity, the 'problem' simply evaporates. (Steedman 1977)

Though he did not put his conclusion in this way, Steedman was essentially saying that Marx cannot be right that labour is the only source of surplus. We are better off to forget the whole question of 'where does the surplus come from?' and instead simply accept that it exists, and analyse capitalism on that basis.

The inconsistencies Steedman establishes[118] undermined Marx's sequence of claims that labour is the only source of value, that value is the only source of profits, and that value determines price. Marx could also provide no reason why capitalism, possible the most internally competitive social system ever, should ultimately behave so co-operatively, with capitalists sharing in total social profit as "just so many stockholders in a stock company in which the shares of profit are uniformly divided per 100".

Thus, though Marx used the labour theory of value to both attack capitalism and predict its downfall, the theory did not even seem to provide a consistent model of capitalism itself – let alone a 'scientific' explanation of why capitalism would wilt and socialism blossom. It appeared that the great revolutionary challenger to capitalism had promised a bang, but delivered a whimper.

The Marxist response

This was no great disappointment to his conservative critics, who happily pointed out the flaws in Marx's logic, and turned to developing economics as we know it today. But devoted Marxists tried valiantly to resurrect Marx's program of 'scientific socialism' by showing that, somehow, at some deep level, Marx's theory of value was internally consistent.

Many years before Steedman turned Sraffa's blowtorch onto Marx's economics, leading Marxist economists had applauded Sraffa's methodical critique of neoclassical economics. However, some of them could also see that Sraffa's dispassionate analysis posed serious problems for the labour theory of value. One of the most thoughtful such responses came from Ronald Meek in his scholarly Studies in the Labour Theory of Value. In a section headed 'From values to prices: was Marx's journey really necessary?', Ronald Meek asked:

> Why did he think that anything had to be 'transformed' in order to arrive at the equilibrium prices characteristic of competitive capitalism? And if something did have to be 'transformed' in order to arrive at them, why did it have to be these mysterious, non-observable, Volume I 'values'?–Personally, although I–am no longer at all religious about such matters, I find myself leaning much more towards the 'neo-Ricardians' than towards their critics. I think that it is

useful to talk in terms of a broad Ricardo-Marx-Sraffa tradition or stream of thought, in which the question of the relation between the social surplus and the rate of profit has always been (and still is) a central theme. (Meek 1973)

In other words, Meek was prepared to abandon the emphasis upon value, and instead develop Marx's analysis of capitalism – minus the insistence that labour is the only source of value, and that value determines profit and prices. Many other scholars followed Meek's lead, and abandoned strict Marxist economics, with its insistence upon value analysis.

However, a minority has persisted, and continue to argue that, somehow, value is an essential part of Marxist analysis. This minority's response to Steedman's critique is best summarised in the title of a paper by Anwar Shaikh: 'Neo-Ricardian economics: a wealth of algebra, a poverty of theory' (Shaikh 1982).

The implication is that, somehow, Marx's philosophy sidesteps the mathematical problems highlighted by Steedman, or it points out a step in the mathematical chain which Steedman missed. To date, no Marxist has been able to put forward an explanation of this rejoinder which has commanded assent from the majority of Marxists: there are almost as many competing ways to try to avoid Steedman's critique as there are Marxist economists. However, they all assent that there is something in Marx's philosophy which counteracts Steedman's mathematical attack.

Over one century after Marx's flawed solution to the transformation problem was first published, and almost a quarter of a century after Steedman's devastating critique, they are still at it. The latest attempts argue that, since Marx's theory was actually dynamic rather than static, the transformation problem should be solvable in a dynamic model.

Nice try, guys, but you really shouldn't bother. The labour theory of value is internally inconsistent, and perhaps even more flawed than conventional economic theory itself. And far from philosophy saving the labour theory of value from mathematical criticism, philosophy provides further compelling reasons for its rejection. One convincing proof of this was given by the Indian economist Arun Bose.

Arun Bose: Marx's 'capital axioms'

Bose was well aware of the criticism levelled at Steedman that his argument, while mathematically impeccable, was somehow philosophically lacking. Though he disparagingly referred to this as "a theological tendency to go so strictly by what Marx said as to adhere to the rule: 'where logic contradicts Marx's words, go by his words'" (Bose 1980), Bose nonetheless tried to avoid this judgment by looking for textual support in Marx. He called his interpretation of Marx the 'capital theory' approach, and argued that:

> as far as logic goes, there are 'two Marxes', the Marx of the 'labour value' approach, and the Marx of the 'capital theory' approach, and that the 'second Marx' should be supported in preference to the first (in scientific discussion). (Bose 1980)[119]

Bose, unlike Steedman, accepted the Marxian position that the concept of value was somehow essential. However, what he argued was that, if value was in some sense the essence of a commodity, then that essence could not be reduced solely to labour. Therefore labour alone was not the essence of value: instead, both labour and commodities were the essence of value. As Bose put it:

> labour is never the only or the main 'source of value' in any system which is defined as capitalist on the basis of a reasonable set of axioms–Labour is not, immediately or ultimately, the only or main source of price, surplus or profit– Labour and commodities are the two sources of wealth, value, price, of surplus value and profit. (Bose 1980)

His logic used a concept we saw earlier in Chapter 6: the reduction of commodity inputs to dated labour.[120]

The manufacture of any commodity requires direct labour, machinery, intermediate goods, and raw materials. All the non-labour inputs had to have been produced at some time in the past: even unprocessed raw materials had to have been previously either mined or harvested. They in turn were made using some direct labour, and other commodity inputs (machinery, intermediate goods, raw materials). These again can be reduced to even earlier dated labour, and other commodity inputs.[121]

This process can go on indefinitely, with each step further reducing the commodity content. But no matter how far back you go, you can never eliminate this commodity residue. If you could, then there would be some commodities that can be created with absolutely no commodity inputs – or in other words, by magic.[122] Therefore if value is the essence of a commodity, then that essence consists of both labour and commodities – it cannot be derived solely from labour.

Bose's conclusion probably helped sway some more Marxists to abandon the faith. But generally, his argument was simply not

acknowledged by Marxist economists. A similar fate has to date befallen the next argument, which establishes that the labour theory of value is inconsistent, not just with mathematical logic, or with any reductionist notion of the commodity, but with Marx's own philosophy.

The origin of surplus value (II)

As noted earlier, most Marxists believe that Marx reached the conclusion that labour was the source of value by a 'negative' proof, which eliminated any other possible contenders. This was true up until 1857, when he developed an alternative, and far superior, 'positive' proof. To understand this proof, we have to delve into Marx's 'dialectical' philosophy.[123]

In brief, dialectics is a philosophy of change. It begins from the proposition that any entity exists in a social environment (see Figure 1). The environment will emphasise some aspect of the entity, and necessarily places less emphasis upon all other aspects of the entity. However, the entity cannot exist without both the foreground aspects (the features the environment emphasises) and background aspects (the ones it neglects). This sets up a tension within the entity, and possibly between the entity and the environment. This tension can transform the nature of the entity, and even the environment itself.

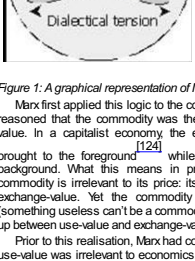


Figure 1: A graphical representation of Marx's dialectics

Marx first applied this logic to the concept of the commodity in 1857. He reasoned that the commodity was the unity of use-value and exchange-value. In a capitalist economy, the exchange-value of a commodity is brought to the foreground[124] while its use-value is pushed into the background. What this means in practice is that the use-value of a commodity is irrelevant to its price: its price is instead determined by its exchange-value. Yet the commodity can't exist without its use-value (something useless can't be a commodity), so that a dynamic tension is set up between use-value and exchange-value in capitalism.

Prior to this realisation, Marx had concurred with Smith and Ricardo that use-value was irrelevant to economics. After it, the concept of use-value, in unison with exchange-value, became a unifying concept for his whole analysis of capitalism.

Marx's first exploration of this concept occurred when he was working on the 'rough draft' of Capital in 1857:

> Is not value to be conceived as the unity of use-value and exchange-value? In and for itself, is value as such the general form, in opposition to use-value and exchange-value as particular forms of it? Does this have significance in economics? (Marx 1857)[125]

The manner in which he first puts the proposition, as questions to himself rather than didactic statements, and especially his comment "Does this have significance in economics?", shows how novel the concept was to him. From this point on, Marx exclusively used this positive methodology, based on a general axiomatic analysis of the commodity, to explain the source of surplus value. Since this point is appreciated by so few Marxists, it is worth citing several of Marx's many pronouncements on this issue.

I noted earlier that Marx mocked Ricardo for not having an explanation of why labour embodied differed from labour commanded. He notes that Smith fell for the fallacy that, under capitalism, a worker should be paid his full product. He continues:

> Ricardo, by contrast, avoids this fallacy, but how? 'The value of labour, and the quantity of commodities which a specific quantity of labour can buy, are not identical.' Why not? 'Because the worker's product–is not = to the worker's pay.' I.e. the identity does not exist, because a difference exists– Value of labour is not identical with wages of labour. Because they are different. Therefore they are not identical. This is a strange logic. There is basically no reason for this other than it is not so in practice. (Marx 1857)

Marx then contrasts his easy ability to derive the source of surplus value with Ricardo's struggles to do the same:

> What the capitalist acquires through exchange is labour capacity; this is the exchange value which he pays for. Living labour is the use-value which this exchange value has for him, and out of this use-value springs the surplus value and the suspension of exchange as such. (Marx 1857)

There are many similar such statements, many of which were written in documents which were either not intended for publication or were never formally completed by Marx. But even in the most well-known passage where Marx derives the source of surplus value, in Capital I, this positive derivation takes precedence over the negative proof.

Marx began Capital by clearing intellectual cobwebs en route to uncovering the source of surplus, criticising explanations based upon unequal exchange or increasing utility through exchange. He then restated the classical axiom that exchange involves the transfer of equivalents, and the conclusion that therefore exchange of itself cannot provide the answer. Yet at the same time circulation based on the exchange of equivalents must be the starting point from which the source of surplus value is deduced. Marx put the dilemma superbly:

> The conversion of money into capital has to be explained on the basis of the laws that regulate the exchange of commodities, in such a way that the starting point is the exchange of equivalents. Our friend, Moneybags, who as yet is only an embryo capitalist, must buy his commodities at their value, must sell them at their value, and yet at the end of the process must withdraw more value from circulation than he threw into it at starting. His development into a full-grown capitalist must take place, both within the sphere of circulation and without it. These are the conditions of the problem. (Marx 1867)

He began the solution of this dilemma with a direct and powerful application of the dialectic of the commodity. If the exchange-value of the commodity cannot be the source of surplus, then the dialectical opposite of value, use-value, is the only possible source:

> The change of value that occurs in the case of money intended to be converted into capital–must–take place in the commodity bought by the first act, M-C, but not in its value, for equivalents are exchanged, and the commodity is paid for at its full value. We are, therefore, forced to the conclusion that the change originates in the use-value, as such, of the commodity, i.e. its consumption. In order to be able to extract value from the consumption of a commodity, our friend, Moneybags, must be so lucky as to find, within the sphere of circulation, in the market, a commodity, whose use-value possesses the peculiar property of being a source of value. (Marx 1867; emphasis added)

Marx then used the quantitative difference between the exchange-value of labour-power and its use-value to uncover the source of surplus value in the transaction between worker and capitalist:

> The past labour that is embodied in the labour power, and the living labour that it can call into action; the daily cost of maintaining it, and its daily expenditure in work, are two totally different things. The former determines the exchange-value of the labour power, the latter is its use-value. The fact that half a [working] day's labour is necessary to keep the labourer alive during 24 hours, does not in any way prevent him from working a whole day. Therefore, the value of labour power, and the value which that labour power creates in the labour process, are two entirely different magnitudes; and this difference of the two values was what the capitalist had in view, when he was purchasing the labour power–What really influenced him was the specific use-value which this commodity possesses of being a source not only of value, but of more value than it has itself. This is the special service that the capitalist expects from labour power, and in this transaction he acts in accordance with the 'eternal laws' of the exchange of commodities. The seller of labour power, like the seller of any other commodity, realizes its exchange-value, and parts with its use-value. (Marx 1867)

The one way in which Marx's 'negative' derivation survived was in the claim that labour-power was the only commodity with the property of being "a source not only of value, but of more value than it has itself". In Capital I, Marx appeared to successfully reach the conclusion that the means of production could not be a source of surplus value. However, he did so by contradicting a basic premise of his 'positive' proof, that the use-value and the exchange-value of a commodity are unrelated. Properly applied, his 'positive proof' contradicts the negative one by showing that all inputs to production are potential sources of surplus-value.

"Guilty of this or that inconsistency because of this or that compromise"

In the course of his attempt to preserve the labour theory of value proposition that labour-power is the only source of surplus value, Marx advanced three propositions which fundamentally contravene his general approach to commodities: that, in the case of the means of production, the purchaser makes use of their exchange-value, not their use-value; that their use-value cannot exceed their exchange-value; and that the use-value of

commodity inputs to production somehow reappears in the use-value of the commodities they help create.

Marx began with the simple assertion that the means of production can transfer no more than their exchange-value to the product. He next attempted to forge an equality between the exchange-value and the use-value of the means of production, by equating the depreciation of a machine to its productive capacity.

> Value exists only in articles of utility–If therefore an article loses its utility, it also loses its value. The reason why means of production do not lose their value, at the same time that they lose their use-value, is this: they lose in the labour process the original form of their use-value, only to assume in the product the form of a new use-value– Hence it follows that in the labour process the means of production transfer their value to the product only so far as along with their use-value they lose also their exchange-value. They give up to the product that value alone which they themselves lose as means of production. (Marx 1867)

Don't worry if you found that paragraph hard to understand: it is replete with erroneous and ambiguous propositions. Firstly, the two final sentences, which appear to link the transfer of value by the machine to its depreciation, are incorrect (see below). Secondly, the statement that the use-value of a machine reappears in the use-value of the product equates the use-value of the machine to the utility enjoyed by the 'consumers' who purchase the goods the machine produces. But the use-value of a machine is specific to the capitalist purchaser of the machine only. By arguing that the use-value of the machine reappears in the product, Marx is in fact contemplating the existence of abstract utility, with the 'usefulness' of the machinery being transmuted into the 'usefulness' of the commodities it produces. If anything, this is neoclassical economics, not Marx.

The ambiguous statement concerns the transfer of value by the means of production. Which of their two 'values' do machines transfer, their exchange-value or their use-value? If Marx meant that they transfer their use-value, then this sentence would be correct in terms of his analysis of commodities. But later he makes it clear that by this expression he meant that the means of production transfer not their use-value (which is the case with a worker) but their exchange-value. In the clearest illustration of the flaw in his logic, he states that over the life of a machine, "its use-value has been completely consumed, and therefore its exchange-value completely transferred to the product" (Marx 1867: 197). This amounts to the assertion that in the case of machinery and raw materials, what is consumed by the purchaser is not their use-value, as with all other commodities, but their exchange-value.

This ambiguity reappears as Marx discusses the example of a machine which only last six days. He first states the correct proposition that the machine transfers its use-value to the product, but then equates this to its exchange-value. He says that if a machine last six days "[t]hen, on the average, it loses each day one sixth of its use-value, and therefore parts with one-sixth of its value to the daily product." Initially he draws the correct if poorly stated inference that "means of production never transfer more value to the product than they themselves lose during the labour-process by the destruction of their own use-value". However the ambiguity between exchange-value and use-value is strong, and his conclusion takes the incorrect fork. Stating his conclusion rather more succinctly than his reasoning, he says

> The maximum loss of value that they [machines] can suffer in the process, is plainly limited by the amount of the original value with which they came into the process, or in other words, by the labour-time necessary for their production–However useful a given kind of raw material, or a machine, or other means of production may be, though it may cost £150–yet it cannot, under any circumstances, add to the value of the product more than £150. (Marx 1867)

Essentially, Marx reached the result that the means of production cannot generate surplus value by confusing depreciation, or the loss of value by a machine, with value creation. The truisms that the maximum amount of value that a machine can lose is its exchange-value, and that a machine's exchange-value will fall to zero only when its use-value has been completely exhausted, were combined to conclude that the value a machine adds in production is equivalent to the exchange-value it loses in depreciation. With the value added by a machine equated to value lost, no net value is transferred to the product, and therefore only labour can be a source of surplus value.

While the argument may appear plausible, in reality it involves a confusion of two distinct attributes of a machine: its cost (exchange-value) and its usefulness (use-value). From a Marxist perspective, depreciation is the writing-off of the original exchange-value of a machine over its productive life. Consequently the maximum depreciation that a machine can suffer is its exchange-value. As it wears out, both its residual value and its usefulness will diminish, and both will terminate at the same time. However, it does not follow that the usefulness (the value-creating capacity) of the machine is equal to its cost (its depreciation). Though a capitalist will

'write off' the latter completely only when the former has been extinguished, the two aspects are nonetheless completely different and unrelated. There is no reason why the value lost by the machine should be equivalent to the value added.

An analogy with labour highlights the fallacy involved in equating these two magnitudes. If workers receive a subsistence wage, and if the working day exhausts the capacity to labour, then it could be argued that in a day a worker 'depreciates' by an amount equivalent to the subsistence wage – the exchange-value of labour-power. However,this depreciation is not the limit of the amount of value that can be added by a worker in a day's labour – the use-value of labour. Value added is unrelated to and greater than value lost; if it were not, there could be no surplus.

But don't take my word for it. Take Marx's.

The origin of surplus value (III)

As noted above, Marx first developed his dialectical analysis of the commodity while working on the rough draft of Capital. He was initially so enthused with this approach that he explored it freely, with almost no regard for how it meshed with his previous analysis. While doing this, he made a statement that correctly applied this new logic and directly contradicted the old, by stating that a machine could add more value than it lost through depreciation.

Table 1 is typical of Marx's standard numerical examples of value productivity. In that table, surplus value is directly proportional to labour-power ('variable capital'), and the value of the total product is the sum of the value of the means of production, plus variable capital, plus surplus value. In this analysis, the contribution of non-labour inputs to the value of output is exactly equal to their depreciation. However, when referring to a similar table shortly after developing his use-value/exchange-value analysis, Marx comments:

> It also has to be postulated (which was not done above) that the use-value of the machine significantly (sic.) greater than its value; i.e. that its devaluation in the service of production is not proportional to its increasing effect on production. (Marx 1857)

There then follows the example shown in Table 7.

Production	Paper	Press	Working Days	Bill	Wage	Surplus	Output
Capital 1	30	30	4		40	10	30
Capital 2	100	60	4		40	13.33	100

Table 8: Marx's example where the use-value of machinery exceeds its depreciation

Both firms employ the same amount of variable capital – 4 days' labour which is paid 40 'thalers' (a unit in the German currency of the time), the value of the labour-power purchased. However, the first firm ('Capital 1'), with older capital, produces surplus value of just 10, while the second, with newer capital, produces a surplus of 131Ú3. The 31Ú3 difference in the surplus they generate is attributable to the difference in their machinery, and the fact that "the use-value of the machine significantly greater than its value; i.e. –its devaluation in the service of production is not proportional to its increasing effect on production".[126]

Marx without the labour theory of value

Marx's dialectical analysis thus contradicts a central tenet of the labour theory of value, that labour is the only source of surplus value. Having reached the conclusion above, Marx suddenly found himself trapped, as he had argued (in his PhD thesis) that Hegel was, in a compromise with his own principles. The principle of the dialectical analysis of the commodity was powerful, and the conclusions that followed logically from it inescapable: the labour theory of value could only be true if the use-value of a machine was exactly equal to its exchange-value, and yet a basic tenet of this analysis was that use-value and exchange-value are incommensurable.[127]

If Marx had followed his newfound logic, the labour theory of value would have been history. But with the labour theory of value gone, so too would be the tendency for the rate of profit to fall, and with it the inevitability of socialism.

The tendency for the rate of profit to fall was predicated upon the propositions that (a) over time, the capital to labour ratio would rise, and that (b) this would cause the rate of profit to fall. But (b) was dependent upon labour being the only source of surplus value, so that a rising capital to labour ratio would mean a falling rate of profit. If surplus could instead be garnered from any input to production, not just labour, then an increase in the capital to labour ratio would have no necessary implications for the rate of profit: it could fall, rise, or stay the same.

With no necessity for the rate of profit to fall, there was similarly no necessity for capitalism to give way to socialism. Yet Marx had prided himself upon being the 'scientific socialist', the one who in contrast to 'utopian socialists' who merely dreamed of a better world, would prove why

socialism had to come about. Now he finds that his new logical tool, which is evidently so superior to his old, challenges the basis of his argument for the inevitability of socialism.

It is little wonder that Marx then tried to find a way to make his new logic appear consistent with the old. By the time of Capital, he had convinced himself that the two were consistent: that the new positive methodology concurred with the old on the issue of the value productivity of machinery. Marx succumbed to the same flaw that (in his PhD thesis) he once noted in Hegel:

> It is conceivable that a philosopher should be guilty of this or that inconsistency because of this or that compromise; he may himself be conscious of it. But what he is not conscious of is that in the last analysis this apparent compromise is made possible by the deficiency of his principles or an inadequate grasp of them. So if a philosopher really has compromised it is the job of his followers to use the inner core of his thought to illuminate his own superficial expression of it. In this way, what is a progress in conscience is also a progress in knowledge. This does not involve putting the conscience of the philosopher under suspicion, but rather construing the essential characteristics of his views, giving them a definite form and meaning, and thus at the same time going beyond them. (Karl Marx 1839: notes to his doctoral dissertation, reprinted in McLellan 1971)

So Marx succeeded in compromising his theory in a way which hid "the deficiency of his principles or an inadequate grasp of them". But 'success' came at a cost. The new logic, of which Marx was so proud, was ignored by his successors. In part, Marx contributed to this by the obfuscation he undertook to make his positive method appear consistent with the old negative one. But I can't detract from the impressive contribution 'Marxists' themselves have made to the misinterpretation of Marx.

The misinterpretation of Marx

Though much of this occurred after his death, Marx had one taste of how his theories would be misinterpreted by friend and foe alike. He wrote a caustic commentary on the German economist Adolph Wagner's gross misinterpretation of his arguments in Capital, yet ironically, Wagner's hostile misinterpretation became the accepted interpretation of Marx by his followers after his death.

Wagner argued that Marx had completely misunderstood the notion of use-value, and that use-value played no part in Marx's analysis. Marx acerbically commented that:

> Rodbertus had written a letter to him–where he, Rodbertus, explains why 'there is only one kind of value', use value–Wagner says: 'This is completely correct, and necessitates an alteration in the customary illogical 'division' of 'value' into use-value and exchange-value'–and this same Wagner places me among the people according to whom 'use-value' is to be completely 'dismissed' 'from science'. (Marx 1879)

Marx then made an emphatic statement of the role that use-value played in his economics:

> All this is 'drivelling'–Only an obscurantist, who has not understood a word of Capital, can conclude: Because Marx, in a note to the first edition of Capital, overthrows all the German professorial twaddle on 'use-value' in general, and refers readers who want to know something about actual use-value to 'commercial guides', – therefore, use-value does not play any role in his work–The obscurantist has overlooked that my analysis of the commodity does not stop at the dual mode in which the commodity is presented, [but] presses forward [so] that–surplus value itself is derived from a 'specific' use-value of labour-power which belongs to it exclusively etc etc., that hence with me use-value plays an important role completely different than [it did] in previous [political] economy. (Marx 1879; emphasis added)

Marx's protestations were to no avail. Despite such a strident statement that use-value was an essential component of his analytic method, and despite the fact that this document was available to and read by early 20th century Marxists, use-value and the 'positive' methodology of which it was an integral part were expunged from mainstream Marxism. Paul Sweezy stated in his influential The Theory of Capitalist Development that

> 'Every commodity,' Marx wrote, 'has a twofold aspect, that of use-value and exchange-value.'–Use-value is an expression of a certain relation between the consumer and the object consumed. Political economy, on the other hand, is a social science of the relations between people. It follows that 'use-value as such lies outside the sphere of investigation of political economy.' (Sweezy 1942, citing Marx 1859)

Yet ironically, the statement Sweezy used to support the notion that use-value plays no role in Marx's analysis was the very one referred to by Marx (in the reference to the "first edition of Capital", by which he meant the 1859 work A Contribution to the Critique of Political Economy) when he labelled

Wagner an "obscurantist". On Marx's own words, therefore, 20th century Marxism has completely misunderstood the philosophical core of Marx's analysis of capitalism.

A poverty of philosophy

Bose's critique and Marx's dialectic of the commodity establish that philosophy can't save the labour theory of value from Steedman's critique. Philosophical analysis strengthens Steedman's case that the labour theory of value is logically flawed.

Instead, mathematics and Marx's philosophy confirm that surplus value – and hence profit – can be generated from any input to production. There is no one source of surplus: Adam Smith's apparently vague musings that animals and machines both contribute to the creation of new value were correct.

Whither Marxism?

Marxist economics is analytically far stronger once it is shorn of the labour theory of value. The use-value/exchange-value methodology, which was applied above only to the question of the source of surplus value, has application to a huge range of issues on which labour theory of value Marxism is either silent or pedestrian (see Groll 1980 and Keen 1993a, 1993b and 2000 for a discussion of some of these). Marxism becomes the pinnacle of classical economics, rather than its dead end.

However, I am as pessimistic about the chances of this 'new, improved Marxism' being adopted by today's Marxists as I am about the chances of neoclassical economists abandoning the concept of equilibrium.

Their resistance, as with neoclassical economists to the critiques outlined in this book, is due in large part to ideology.

The advantage Marxists have over economists is that at least they are upfront about having an ideology. Marxists are as consciously committed to the belief that capitalism should give way to a socialism as economists are to the often unconscious belief that, if only we could rid ourselves of government intervention in the market, we would currently reside in the best of all possible worlds.

The tendency for the rate of profit to fall is crucial to this belief in the inevitability of socialism, and it is one of the many concepts that evaporate once the labour theory of value is expunged. Marxist economists are likely to continue to cling to the labour theory of value, to hang on to the faith, in preference to embracing logic.

If my pessimism is well founded, then Marxist economics will continue its self-absorbed and impossible quest for a solution to the transformation problem, and will remain irrelevant to the future development of economics.

However, labour theory of value Marxism will continue to be the ideology of choice of the Left, particularly in the Third World. The argument that labour is the only source of profit, and that capitalism is thus based upon the exploitation of the worker, is a simple, compelling analysis to the downtrodden in our obscenely unequal world. A spectre may no longer be haunting Europe, but Marxism will continue to be the banner of the dispossessed for many a year to come.

However, if non-neoclassical and non-Marxist economists can ignore the hullabaloo generated by the remaining band of adherents to the labour theory of value, and instead extract from Marx his rich philosophical foundation for the analysis of capitalism, then Marx's dialectical theory of value may yet play a role in the reform of economic theory. At present, however, the various non-neoclassical schools of thought have no coherent theory of value as an alternative to the neoclassical school's flawed subjective theory of value. But even though they lack the central organising concept of a theory of value, these alternative schools of thought contain the promise of an economic theory that may actually be relevant to the analysis and management of a capitalist economy.

14. There are alternatives[128]

Why there is still hope for a better economics

Maggie Thatcher's second best-known comment, in defence of following monetarist economic policies, was "There is NO alternative." A similar attitude pervades economics: if not neoclassical economics, then what?

In fact, there are many alternative schools of thought within economics. In addition to Marxian economics, the main alternatives are:

- Austrian economics, which shares many of the features in common of neoclassical economics, save a slavish devotion to the concept of equilibrium;
- Post-Keynesian economics, which is highly critical of neoclassical economics, emphasises the fundamental importance of uncertainty, and bases itself upon the theories of Keynes and Kalecki;
- Sraffian economics, based on Sraffa's concept of the production of commodities by means of commodities;
- Complexity theory/ecophysics, which applies concepts from nonlinear dynamics, chaos theory and physics to economic issues; and
- Evolutionary economics, which treats the economy as an evolving system along the lines of Darwin's theory of evolution.

None of these is at present strong enough or complete enough to declare itself a contender for the title of 'the' economic theory of the 21st century. However, they all have strengths in areas where neoclassical economics is fundamentally flawed, and there is also a substantial degree of overlap and cross-fertilisation between schools. It is possible that this century could finally see the development of a dominant economic theory which actually has some relevance to the dynamics of a modern capitalist economy.

I would probably be regarded as partisan to the Post-Keynesian approach. However, I can see varying degrees of merit in all five schools of thought, and I can imagine that a 21st century economics could be a m–lange of all five.

In this chapter I will give a brief overview of each school, emphasising the ways in which they are superior to neoclassical economics, but also noting when they share its weaknesses, or have problems of their own. This will necessarily be an inadequate survey – doing a proper survey would necessitate another book. But as I commented earlier, it is essential to at least outline the alternatives, to debunk the myth that there is no alternative.

Austrian economics

The Austrian school (so-called because its main early protagonists – Menger, Hayek and von Mises – were Austrian, though it is now mainly an American tradition) is a close relative of mainstream economics. It developed at much the same time, shared the same intellectual parents, and is comfortable with virtually every aspect of neoclassical economics save one: its obsession with equilibrium. This one divergence results in a theory which is markedly different from its dominant but wayward cousin.

Strengths

Far from arguing that capitalism is the best social system because of the conditions which pertain in equilibrium, Austrian economists argue that capitalism is the best social system because of how it responds to disequilibrium.

The Austrians make the sensible observation that equilibrium is an intellectual abstraction which is unlikely ever to occur in the real world. All real-world economic situations will thus be disequilibrium ones, some of which enable entrepreneurs to make above-normal profits. By seeking out these profit opportunities, entrepreneurs make capitalism a dynamic, adaptive social system.

The Austrians therefore have an evolutionary perspective on capitalism, and argue that capitalism is evolutionarily superior to other social systems, such as feudalism and socialism.

The Austrians emphasise the importance of uncertainty in analysing capitalism, whereas neoclassical economists, as we have seen, either ignore uncertainty, or trivialise it by equating it to probabilistic risk. This again gives Austrians an ideological reason to prefer capitalism to any other social system, since they argue that the disaggregated nature of capitalist society makes it more adaptable to uncertainty than other, more centralised systems.

The entrepreneur is the key actor in the Austrian vision of capitalism. It is the entrepreneur who attempts to profit from disequilibrium situations, thus innovating and adding to the diversity and strength of the capitalist system. The entrepreneurs are those who boldly act in the face of uncertainty, and though many will fail, some will succeed – thus strengthening the economic system via an evolutionary process.

The Austrians thus demonstrate that the economic fixation with equilibrium is unnecessary: it is possible to be an ideological supporter of capitalism even if you believe that equilibrium is irrelevant.

However, as a near relative of neoclassical economics, this school shares a number of its disabilities.

Weaknesses

Firstly, the Austrians accept the economic argument that production is characterised by diminishing returns. As a corollary of this, they also accept the marginal productivity theory of income distribution – though they temper this by arguing that disequilibrium allows for entrepreneurs to make super-normal profits.

As was shown in chapters 3, 5 and 6, these economic notions are fundamentally unsound. To the extent that Austrian economics relies upon these same concepts, it is also unsound.

A simple illustration of this arises from the Austrian theory of production. The economic model argues that an increase in the quantity of a factor of production – such as capital – will decrease its marginal product, and thus reduce its income.

The Austrians instead argue that a cheapening of capital – via a fall in the rate of interest – will lead to a more 'roundabout' approach to production, meaning that less direct labour and more indirect capital will be applied to its production.

Sraffa's critique of the neoclassical theory of production, detailed in Chapter 6, is equally applicable to this Austrian theory. By providing a way to measure capital inputs in terms of wage units, Sraffa showed that the economic concept of a quantity of capital was dependent on the rate of profit: the same logic shows that it is impossible to define one way of producing a commodity as 'more roundabout' than another independently of the rate of profit.

Consider two ways of making wine: process A which involves the application of 1 wage unit now, 8 units last year, and 1 unit 8 years earlier; and process B, which involves 1 unit now and 1 unit 20 years ago. At a low rate of profit, process A might be more roundabout than process B; at a higher rate of profit, the order could reverse; and it could reverse again for a higher rate of profit. Therefore, the Austrian notion of roundaboutness is as internally inconsistent as the neoclassical concept of the marginal productivity of capital.

Secondly, even more so than conventional economics, Austrian economics has a faith in the self-adjusting properties of the capitalist economy, with Say's Law providing much of that confidence. As was argued in Chapter 9, Say's Law is invalid in a production economy with growth.

Thirdly, while it is in general an evolutionary approach to economics, at least one branch of Austrian economics, associated with Murray Rothbard, has a quite non-evolutionary attitude towards both the existence of the State, and the role of money. The market economy may have evolved, but it seems the State was simply imposed from outside as an alien artefact upon our landscape. This is certainly one way to consider the growth of the welfare state; but an equally tenable argument is that the welfare state evolved as a response to the failures of the pure market system during the Great Depression.

Similarly, while they believe that the money supply should be determined endogenously – by either handing over money creation to private banks, or by returning to the gold standard – they argue that the current system of State money means that the money supply is entirely exogenous, and under the control of the State authorities. They then attribute much of the cyclical behaviour of the economy to government meddling with the money supply and the rate of interest.

The Post-Keynesian school, on the other hand, argues that though it may appear that the State controls the money supply, the complex chain of causation in the finance sector actually works backwards. Rather than the State directly controlling the money supply via its control over the issue of new currency and the extent to which it lets banks leverage their holdings of currency, private banks and other credit-generating institutions largely force the State's hand. Thus the money supply is largely endogenously determined by the market economy, rather than imposed upon it exogenously by the State.

The empirical record certainly supports Post-Keynesians rather than Austrians on this point. Statistical evidence about the leads and lags between the State-determined component of money supply and broad credit show that the latter 'leads' the former (Kydland and Prescott 1990). If the Austrians were correct, State money creation would instead precede private credit creation.

Maggie Thatcher's embrace of monetarism also provides an evocative counter-example. Despite her toughness, and her adherence to Milton Friedman's mantra that controlling the money supply would control inflation, even Thatcher's England was forced to abandon monetary targeting – setting some goal for the rate of growth of the money supply in order to force down the rate of inflation – because it could never meet the targets. If 'The

Iron Lady' couldn't control the money supply, then no-one could: evidence enough that the Post-Keynesians are closer to the mark than the Austrians.

This non-evolutionary weakness in Austrian economics is a sign of a wider problem. The philosopher Chris Sciabarra, a specialist on the Austrian school and Ayn Rand, identifies an inconsistency between Hayek's notion of 'spontaneous order' – which corresponds to evolutionary development – and 'designed order' – where change is imposed from outside the market by the State. While such a distinction makes for good polemic writing against State intervention, it ignores the extent to which the State's own behaviour might be reactive to the market, and thus, to some extent, also a form of spontaneous order. As Sciabarra puts it:

> There are more fundamental problems with Hayek's social theory. By positing such a sharp distinction between spontaneous order and designed order, Hayek has not provided us with any explanation of the emergence of those institutions which are agents of constructivism [designed order]–To what extent is the state itself a spontaneous, emergent product of social evolution? To what extent does the state define the parameters of the extended order which Hayek celebrates? What are the actual interrelationships between the spontaneous order of the market and the designed institutions of the state? The reader of Hayek's works will strain to find developed answers to any of these important questions. (Sciabarra 1995)

Finally, though Austrians eschew equilibrium analysis, and regard it as an unattainable state, their preference for capitalism as a social system is partly dependent on the belief that it will remain close to equilibrium. If, instead, capitalism is endogenously unstable, then it may remain substantially distant from equilibrium situations all the time. This weakens Austrian economics, to the extent that its support for capitalism emanates from conditions which are assumed to apply in equilibrium.

The Austrian scorecard

Overall, I regard the Austrian school as too close to its neoclassical cousin to make a major contribution to a reformed economics. However, it does have some contributions to make, and for ideological reasons it is likely to be far stronger in the future – regardless of what I or other non-orthodox economists might think of it.

The Austrian emphasis upon innovation, and the role of the entrepreneur, are valid concepts which capture the way in which a market economy adapts. This aspect of capitalism is to some extent under-rated by the other non-neoclassical schools, except for evolutionary economics. This aspect of Austrian thought could be valuable to a reformed 21st century economics.

However, it is far more likely that the 'pure and simple' Rothbardian stream of Austrian economics will play a large role in 21st century economics. This is because the Rothbardian approach provides an alternative way to ideologically support a capitalist economy as the best possible social system, whereas all other non-orthodox schools are to some degree critical of the concept of unfettered capitalism. If neoclassical economics becomes untenable for any reason, the Austrians are well placed to provide an alternative religion for believers in the primacy of the market over all other forms of social organisation.

The one barrier which stands in the way of today's neoclassical economist transmuting into tomorrow's Austrian is the Austrian insistence that there is little, if any, role for mathematics in economic analysis. Because the Austrians believe that all real-world data is generated in a situation of disequilibrium, and because they take seriously the aggregation problems noted in chapters 2 and 4, Austrians deny that mathematical aggregate analysis has any validity. Faced with a choice between ideology and their beloved equilibrium mathematics, most economists would probably prefer to keep the mathematics. The one way out for neoclassicals would be to embrace the Austrian celebration of capitalism as a dynamic, disequilibrium system, and then model this using chaos and complexity theory. But this leads to the dilemma that such models almost always display 'far from equilibrium' behaviour, which undermines the validity of beliefs about capitalism and welfare that depend on the economy not straying too far from equilibrium.

Post-Keynesian economics

This school of thought developed in reaction to the 'bastardisation' of Keynes's economics in the so-called Keynesian-neoclassical synthesis. Regarding themselves as the true carriers of Keynes's message, they emphasised the importance of uncertainty in economic analysis, and the profound difference between the monetary economy in which we live, and the barter economy which neoclassical economics regards as an adequate proxy for the real world. As Arestis, Dunn and Sawyer put it, the main unifying themes in Post-Keynesian economics are "a concern for history, uncertainty, distributional issues, and the importance of political and economic institutions in determining the level of activity in an economy" (Arestis, Dunn & Sawyer 1999).

Strengths

The emphasis upon uncertainty as a fundamental aspect of the real world – one which cannot be approximated by risk – makes the Post-Keynesian approach to economics far more realistic than the neoclassical.

Realism, in fact, is a central methodological emphasis of this school. Though there is no agreed Post-Keynesian methodology to rival the hedonistic calculus of the neoclassicals, Post-Keynesians are united by their belief that an economic model has to be realistic.

One essential aspect of this is the insistence that a monetary economy is fundamentally different from one in which commodities simply exchange against other commodities. The issues of credit creation, the nature of money, the role of debt, etc., are far more pressing to Post-Keynesians than they are to neoclassicals.

Macroeconomics is also a far more important concern. In fact, Post Keynesians reverse the neoclassical pecking order, to argue that whatever microeconomics is developed must be consistent with the observed behaviour of the macroeconomy. A microeconomic model which is inconsistent with such things as business cycles, sustained unemployment, commonplace excess capacity, and the importance of credit, is to Post-Keynesians an invalid model.

Their preferred model of the firm emphasises monopoly and quasi-monopoly power, decreasing costs of production with increased scale, markup pricing, the competitive need to sustain excess capacity, and the link between macroeconomic conditions and the firm's investment decisions.

Post-Keynesians are also not hung up on the need to show whether capitalism is a better or worse social system than any other. They are relatively agnostic on the question of what might constitute a better society.

This comparative independence from ideology means that Post-Keynesians do not feel compelled, as neoclassicals do, to show that capitalism generates the best welfare outcome for the majority of the people. They are therefore much more comfortable with acknowledging the existence of social class in their models – something which leading neoclassical economists admit they might have to consider, given the failure of their individualistic approach to explain human behaviour.

Weaknesses

One of this school's great strengths is also a weakness. Unlike the neoclassical or Marxian schools, they do not have a 'theory of value' – they have nothing to compare to the theory of utility maximisation, or even the labour theory of value.

This is certainly a strength when one considers how these theories of value have led these rival schools up intellectual garden paths. However, at the same time it means that Post-Keynesians lack a methodological consistency: they are more united by what they oppose than by what they have in common – though there are many common threads.

This lack of a theory of value makes it difficult for Post Keynesians to explain why their approach is superior to fledgling students of economics, who have yet to confront any of the intellectual conundrums which afflict neoclassical economics (and they also have difficulty communicating with radical students who are attracted to Marxism). One must normally become disenchanted with mainstream or Marxian economics before one can become a Post-Keynesian. That is perhaps too tortuous a path to rely upon, if this school ever hopes to gain the ascendancy in economics.

A final problem is that, despite their rejection of neoclassical economics, they tend to also use static logic in their analysis – even though their building blocks might be, for instance, markup pricing rather than 'marginal cost equals marginal revenue'. This lack of appreciation of how different dynamic analysis is from static is not universal amongst Post-Keynesian authors, but it is widespread enough to be a problem. However, it must be said that younger members of the Post-Keynesian school are much more comfortable with dynamic analysis than are some of its older members.

The Post-Keynesian scorecard

Despite the lack of an agreed methodological foundation, the Post-Keynesians are easily the most coherent alternative school of economic thought today. They are also likely to gain substantial credence in the event of a financial crisis, given their explicitly monetary approach to economics.

Sraffian economics

No prizes for guessing which economist provided the major inspiration to this particular group of economists. Sraffa's Production of Commodities by Means of Commodities became the icon, for these economists. As well as applying it to critique other schools – notably neoclassical economics and Marxism – they attempted to turn it into a means to analyse the real economy.

Strengths

There is no doubt that Sraffa's analysis constituted the most detailed and careful analysis of the mechanics of production in the history of economics. Not for him any simple abstractions, such as the neoclassicals' 'factors of production', or even Marx's 'industry sectors': his model analysed the interrelations of production at the level of the individual commodity.

This study turned up many subtleties that escaped other schools of economics: the dependence of the 'quantity of capital' on the profit rate, rather than vice versa, the phenomenon of reswitching, etc. No other school of economics matches the Sraffians on this insistence of assumption-free rigour. Well, almost assumption-free rigour.

Weaknesses

The one assumption Sraffians do make is that the economy can be analysed using static tools. As a result, even though the proper treatment of time was an essential component of Sraffa's critique of neoclassical economics, modern Sraffian economics makes no use of time or dynamics. Ian Steedman gave the pithiest explanation of why Sraffians analysis ignores dynamics. It is because

> 'static' analysis does not 'ignore' time. To the contrary, that analysis allows enough time for changes in prime costs, markups, etc., to have their full effects. (Steedman 1992)

This proposition can only be true if the long run position of an economy is an equilibrium one: if, in other words, the economy has just one equilibrium, and it is stable. As Chapter 8 showed, this is highly unlikely to be the case. A market economy is likely to have multiple equilibria, and they are all likely to be unstable. The Sraffian position is thus ignorant of modern dynamic analysis.

Sraffians also have one other flaw: they pay too little attention to Piero Sraffa.

Some Post-Keynesians are fond of pointing out how pedantic Sraffa was, and therefore how important was the subtitle to his magnum opus. Sitting beneath the title of The Production of Commodities by Means of Commodities was the caveat 'Prelude to a critique of economic theory'.

In other words, these economists argue that Sraffa's method was intended solely to provide a means to critique other economic theories: it was never meant to provide a basis for an economic theory in itself.

Sraffa's 1926 paper provides support for this position. When discussing how the firm should be modelled, Sraffa put great stress upon the issues of importance to 'business men': the necessity and expense of marketing a non-homogeneous product to a market of non-homogeneous consumers, the cost and dangers of credit as a major force limiting firm size, etc. The concepts Sraffa discusses here can be considered only with extreme difficulty in the framework of his 1960 book (check the web link Alternatives/Sraffa for a relevant extract from Sraffa's 1926 paper).

The Sraffian scorecard

The argument that their analysis wrongly ignores dynamics has recently been put to Sraffians (Keen 1998), and the jury is still out on how they will respond. If they succeed in melding a dynamic approach to economics with their multi-sectoral rigour, they could be a major force in future economic analysis.

Complexity theory/Econophysics

Complexity theory is not so much a school of thought in economics as a group of economists who apply what is popularly known as 'chaos theory' to economic issues.

The concept of chaos itself was first discovered in 1899 by the French mathematician Henri Poincar–. However, knowledge of it languished until the mid-1960s because it could not be fully explored until after the invention of computers. Chaotic models of necessity cannot be understood simply by writing down the equations which represent them: instead, they must be simulated, and their properties analysed numerically. This was simply not possible before the advent of computers.

An essential aspect of complexity is the existence of nonlinear relationships between elements of a system, and the apparent ability of complex systems to 'self-organise'. The Lorenz model, noted in Chapter 8, has both these attributes: the nonlinear relations between displacement and temperature lead to behaviour which on the surface is 'chaotic', but behind which lies the beautiful organising force of Lorenz's 'strange attractor'. Complexity theorists argue that the economy demonstrates similar attributes, and these are what give rise to the cycles which are a self-evident aspect of real-world economies.

Strengths

It is extremely difficult to work in complexity theory and not understand dynamics. Though some neoclassical dabblers occasionally attempt to use equilibrium thinking in so-called chaotic models, in the main, practitioners in this camp are extremely well versed on dynamics.

They are also normally very competent in mathematics; far more so, not only than other alternative schools of economics, but also than the neoclassicals. Many complexity theorists in economics started out doing PhDs in physics, biology, or mathematics itself, and later delved into economics out of curiosity.

This alone has meant that complexity theorists have had a significant impact upon the profession. While they rarely indulge in direct attacks upon neoclassical economics per se, neoclassical economists are aware that they are quite capable of doing so if provoked. This technical superiority over neoclassical economists has taken the mathematical 'big stick' out of the hands of neoclassicals.

Weaknesses

Though many complexity economists are inclined to a Post-Keynesian perspective on economics, in general they lack a full appreciation of the history of economic thought. For this reason, they will often generate models which combine incompatible streams in economics. Concepts such as IS-LM and rational expectations often crop up in complexity models of the economy, with the authors rarely being aware of the origins of these 'tools'.

Econophysics has developed a very rich and empirically based analysis of financial markets to date, and their statistical analysis here—involving concepts like Power Law distributions and Tsallis-statistics—is far more accurate than neoclassical models. Success here has led to neglect of the "econo" part of the developing discipline's name: at present it could more accurately be called "finaphysics" than "econophysics".

The Complexity scorecard

Complexity theory and econophysics are amongst the 'glamour' areas of science in general today, and this affects economics, even given its relative isolation from the scientific mainstream. The techniques which complexity modellers in economics employ are thus 'refertilising' economics with concepts from other disciplines. The economic fixation upon equilibrium appears quaint to these mathematically literate economists, and this alone may significantly undermine the hold which static thinking has on economics.

If statics were to die, then inevitably so too would neoclassical economics, since its way of thinking is unsustainable in dynamics. So complexity theory may be a harbinger of real change in economics, after sixty years of effective ossification.

Evolutionary economics

Evolutionary economics draws its inspiration from the theory of evolution. In this, it has much in common with the majority of physical sciences, which in recent years have started to apply the concept of evolution – so much so that it has been proposed that Darwinism is the 'universal' basis of science (Nightingale 2000).[129]

In all sciences, the basic building blocks of the evolutionary way of thinking are diversity, the environment, and adaptation. Diversity gives a range of possible 'solutions' to the challenges thrown up by the environment. The environment interacts with these diverse forms to favour some over others – and the environment itself may be altered by feedback between it and these newly emergent species (Levins & Lewontin 1986). Adaptation occurs at the systemic level through the differential survival of some of these diverse forms: while some die out, others prosper, and thus their characteristics are passed on more strongly to subsequent generations.

The economic equivalents of diversity are the heterogeneity of consumers, and the variety of commodities; the equivalents of adaptation are new product development, and the consequent endogenous alteration of consumer tastes; the equivalent of the environment is the economy itself, which is endogenously created by the actions of myriad individuals, social groups and corporations.

Strengths

It is undeniable that the economy is an evolutionary system – with the one embellishment that change in economics is often purposeful, as opposed to the random nature of variation in the environment (though of course, purposive change can fail to achieve its intended ends).[130] This self-evident fact was the basis for Thorstein Veblen's query, over a century ago, of 'Why is economics not an evolutionary science?'

Manifestly it should be, and this alone should be a major factor in the rise of evolutionary thinking in economics.

Weaknesses

One problem with evolutionary systems is that, effectively, everything can change. Economists, on the other hand, have been wedded to the notion of 'ceteris paribus' ('all other things remaining equal') as a way of being able to impose some order on the apparent chaos of the market.

Ceteris paribus is of course an illusion, but the illusion often seems preferable to reality when it appears that fully acknowledging reality forces one to abandon structure.

This, of course, is not correct: evolutionary modelling still has structure, as shown by the advances made in genetics and many other areas where evolutionary thinking rules. However, economists are thrown back upon analogy here, since in economic systems there is no comparable entity to the gene, nor to the processes of biological interaction.

The difficulty evolutionary economists face is developing analytic tools which are consistent with evolution, and yet which still enable meaningful statements to be made about economic issues.

The Evolutionary scorecard

Evolutionary economics is still in its infancy, and a lot of its time is spent defining basic philosophical concepts at one extreme, and developing computer-based evolutionary economic models at the other. If it can coalesce into a coherent school of thought with effective analytic tools, then it could at last make economics what, one century ago, Veblen knew it should be: an evolutionary science.

W(h)ither economics?

This is far from the first book to attack the validity of economics, and it is unlikely to be the last. As Kirman commented, economic theory has seen off many attacks, not because it has been strong enough to withstand them, but because it has been strong enough to ignore them.

Part of that strength has come from the irrelevance of economics. You don't need an accurate theory of economics to build an economy in the same sense that you need an accurate theory of propulsion to build a rocket. The market economy began its evolution long before the term 'economics' was ever coined, and it will doubtless continue to evolve regardless of whether the dominant economic theory is valid. Therefore, so long as the economy itself has some underlying strength, it is a moot point as to whether any challenge to economic orthodoxy will succeed.

However, while to some extent irrelevant, economics is not 'mostly harmless'. The false confidence it has engendered in the stability of the market economy has encouraged policy-makers to dismantle some of the institutions which initially evolved to try to keep its instability within limits. 'Economic reform', undertaken in the belief that it will make society function better, has instead made modern capitalism a poorer social system: more unequal, more fragile, more unstable. And in some instances, as in Russia, a naive faith in economic theory has led to outcomes which, had they been inflicted by weapons rather than by policy, would have led their perpetrators to the International Court of Justice.

But even such a large-scale failure as Russia seems to have little impact upon the development of economic theory. For economics to change, it appears that things have to 'go wrong' on a global scale, in ways which the prevailing theory believed was impossible. There have been two such periods this century.

The first and most severe was the Great Depression, and in that calamity, Keynes turned economic theory upside down. However, Keynes's insights were rapidly emasculated, as Chapter 9 showed. 'Keynesian economics' became dominant, but it certainly was not the economics of Keynes.

The second was the 'stagflationary crisis' – the coincidence of low growth, rising unemployment and high inflation during the 1970s. That crisis led to the final overthrow of the emasculated creature that Keynesian economics had become, and its replacement by an economic orthodoxy which was even more virile than that against which Keynes had railed.

One step forward and two steps back – with the first step backwards being taken when the economy was doing well, in the aftermath of the Depression and WWII and hence when the ramblings of economists could comfortably be ignored.

That historical record is both comforting and disturbing. Change is possible in economics, but normally only when the fabric of society itself seems threatened; and change without crisis can involve the forgetting of recent advances.

It is possible, therefore, that economic theory may continue to function mainly as a surrogate ideology for the market economy, right up until the day, in some distant future, when society evolves into something so profoundly different that it no longer warrants the moniker 'capitalism'.

I hope, however, that events follow a different chain. I am not wishing an economic crisis upon the modern world – instead, I think one has been well and truly put in train by the cumulative processes described in chapters 10 and 11. If that crisis eventuates – one which neoclassical economic theory argues is not possible – then economics will once again come under close and critical scrutiny.

This time, the chances are much better that something new and indigestibly different from the prevailing wisdom will emanate from the crisis. As this book has shown, critical economists are much more aware of

the flaws in conventional economics than they were during Keynes's day, non-orthodox analysis is much more fully developed, and advances in many other fields of science are there for the taking, if economics can be persuaded – by force of circumstance – to abandon its obsession with equilibrium.

The first factor should mean that the lines will be much more clearly drawn between the old orthodoxy and the new. The latter two should mean that the techniques of the old orthodoxy will look pass–, rather than stimulating, to a new generation of economists schooled in complexity and evolutionary theory.

But ultimately, schooling is both the answer and the problem. If a new economics is to evolve, then it must do so in an extremely hostile environment – the academic journals and academic departments of Economics and Finance, where neoclassic orthodoxy has for so long held sway.[131] The nurturing of a new way of thinking about economics could largely be left in the hands of those who have shown themselves incapable of escaping from a 19th century perspective.

The are two possible palliatives against that danger. The first is the development, by non-orthodox economists, of a vibrant alternative approach to analysing the economy which is founded in realism, rather than idealism. Such a development would show that there is an alternative to thinking about the economy in a neoclassical way, and offer future students of economics a new and hopefully exciting research program to which they can contribute.

The second is an informed and vigilant public. If you have struggled to the end of this book, then you now have a very strong grasp on the problems in conventional economic thought, and the need for alternative approaches to economics. Depending on your situation, you can use this knowledge as a lever in all sorts of ways.

If you are or you advise a person in authority in the private or public sectors, you should know now not to take the advice of economists on faith. They have received far too easy a ride as the accepted vessels of economic knowledge. Ask a few enquiring questions, and see whether those vessels ring hollow. When the time comes to appoint advisers on economic matters, quiz the applicants for their breadth of appreciation of alternative ways to 'think economically', and look for the heterodox thinker rather than just the econometric technician.

If you are a parent with a child who is about to undertake an economics or business degree, then you're in a position to pressure potential schools to take a pluralist approach to education in economics. A quick glance through course structure booklets and subject outlines should be enough to confirm what approach they take at present.

If you are a student now? Well, your position is somewhat compromised: you have to pass exams, after all! I hope that, after reading this book, you will be better equipped to do that. But you are also equipped to 'disturb the equilibrium' of both your fellow students and your teachers, if they are themselves ignorant of the issues raised in this book.

You have a voice, which has been perhaps been quiescent on matters economic because you have in the past deferred to the authority of the economist. There is no reason to remain quiet.

I commented at the beginning of this book that economics was too important to leave to the economists. I end on the same note.

References

Preface

Keynes, J.M. (1936). *The General Theory of Employment, Interest and Money*, Macmillan, London.

Lancaster, K. & Lipsey, R.G., (1956). 'The general theory of the second best', *Review of Economic Studies*, 24: 11–32. Published in K. Lancaster (1996), *Trade, Markets and Welfare*, Edward Elgar, Cheltenham.

No more Mr Nice Guy

Leijonhufvud, Axel, (1973). 'Life among the Econ', *Western Economic Journal*, 11: 327–337. Republished in C. P. Clotfelter (ed.), (1996), *On the Third Hand: Humor in the Dismal Science*, University of Michigan Press, Ann Arbor: 24–35.

McDonough, T. & Eisenhauer, J., (1995). 'Sir Robert Giffen and the great potato famine: a discussion of the role of a legend in neoclassical economics', *Journal of Economic Issues*, 29: 747–759.

Robbins, L., (1932 [1952]). *An Essay on the Nature and Significance of Economic Science*, 2nd edition, Macmillan, London.

Stiglitz, J., (1998). 'The confidence game: how Washington worsened Asia's crash', *The New Republic Online*, 9 September.

Stiglitz, J., (1993). 'Post Walrasian and post Marxian economics', *Journal of Economic Perspectives*, 7 (1): 109–114.

Stiglitz, J., (2000). 'What I learned at the world economic crisis', *The New Republic*, April 17–24: 56–60.

The calculus of hedonism

Barnett, W.A., (1979). 'Theoretical foundations for the Rotterdam model', *Review of Economic Studies*, 46: 109–130.

Battalio, R.C., Kagel, J.H., Winkler, R.C., Fisher, E.B., Bassmann, R.L. & Krasner, L., (1977). 'A test of consumer demand theory using observations of individual consumer purchases', *Western Economic Journal* 11: 411–428.

Bentham, J., (1780 [1948]). *The Principles of Morals and Legislation*, Hafner Press, New York.

Bowles, H. Gintis, S., (1993). 'The revenge of Homo Economicus: contested exchange and the revival of political economy', *Journal of Economic Perspectives*, 7 (1): 83–102.

Carter, J.R. & Irons, M.D., (1991). 'Are economists different, and if so, why?', *Journal of Economic Perspectives*, 5 (2): 171–177.

Chipman, J.S., (1974). 'Homothetic preferences and aggregation', *Journal of Economic Theory*, 8: 26–38.

Debreu, G., (1970). 'Economies with a finite set of equilibria', *Econometrica*, 38: 387–392.

Debreu, G., (1974). 'Excess demand functions', *Journal of Mathematical Economics*, 1: 15–21.

Dierker, E., (1972). 'Two remarks on the number of equilibria of an economy', *Econometrica*, 40: 951–953.

Dierker, E., (1977). 'Generalised Slutsky conditions for aggregate consumer demand functions', *Journal of Economic Theory*, 15: 353–362.

Earl, P. E., (1995). *Microeconomics for Business and Marketing*, Edward Elgar, Cheltenham.

Frank, R.H., Gilovich, T. Regan, D.T., (1993). 'Does studying economics inhibit cooperation?', *Journal of Economic Perspectives*, 7 (2): 159–171.

Frank, R.H., Gilovich, T. & Regan, D.T., (1996). 'Do economists make bad citizens?', *Journal of Economic Perspectives*, 10 (1): 187–192.

Gorman, W.M., (1953). 'Community preference fields', *Econometrica*, 21: 63–80.

Hirshleifer, J., (1993). 'The dark side of the force', *Economic Inquiry*, 32: 1–10.

Kirman, A., (1989). 'The intrinsic limits of modern economic theory: the emperor has no clothes', *Economic Journal*, 99: s126–s139.

Kreps, D.M, (1990). *A Course in Microeconomic Theory*, Princeton University Press, New Jersey.

Lavoie, M, (1994). 'A Post Keynesian approach to consumer choice', *Journal of Post Keynesian Economics*, 16: 539–562.

Marwell, G. & Ames, R., (1981). 'Economists free-ride, does anyone else?', *Journal of Public Economics*, 15 (3): 295–310.

Mantel, R.R., (1974). 'On the characterisation of aggregate excess demand', *Journal of Economic Theory*, 7: 348–353.

Mantel, R.R., (1976). 'Homothetic preferences and community excess

demand functions', *Journal of Economic Theory*, 12: 197–201.

Marks, R., (2000). 'Evolved perception and the validation of simulation models', in W.A. Barnett, C. Chiarella, S. Keen, R. Marks, H. Schnabl (eds), *Commerce, Complexity and Evolution*, Cambridge University Press, New York.

Mas-Colell, A, (1977). 'On the equilibrium price set of an exchange economy', *Journal of Mathematical Economics*, 4: 117–126.

McFadden, D., Mas-Colell, A, & Mantel, R.R., (1974). 'A characterisation of community excess demand functions', *Journal of Economic Theory*, 9: 361–374.

Oda, S. H., Miura, K., Ueda, K. and Baba, Y, (2000). 'The application of cellular automata and agent models to network externalities in consumers' theory: a generalization of life game', in W.A. Barnett, C. Chiarella, S. Keen, R. Marks, H. Schnabl (eds), *Commerce, Complexity and Evolution*, Cambridge University Press, New York.

Roth, T.P., (1989). *The Present State of Consumer Theory*, University Press of America, Lanham.

Samuelson, P.A, (1948). *Foundations of Economic Analysis*, Harvard University Press, Cambridge MA.

Sent, E.M., (1997). 'Sargent versus Simon: bounded rationality unbound', *Cambridge Journal of Economics* 21: 323–328.

Shafer, W.J., (1977). 'Revealed preference and aggregation', *Econometrica*, 45: 1173–1182.

Shafer, W & Sonnenschein, H., (1982). 'Market demand and excess demand functions', in K.J. Arrow, and M. D. Intriligator (eds), *Handbook of Mathematical Economics* (Vol. II), North-Holland, Amsterdam.

Sippel, R., (1997). 'Experiment on the pure theory of consumer's behaviour', *Economic Journal* 107: 1431–1444.

Sonnenschein, H., (1972). 'Market excess demand functions', *Econometrica*, 40: 549–563.

Sonnenschein, H., (1973a). 'Do Walras' identity and continuity characterise the class of community excess demand functions', *Journal of Economic Theory*, 6: 345–354.

Sonnenschein, H., (1973b). 'The utility hypothesis and demand theory', *Western Economic Journal*, 11: 404–410.

Stiglitz, J., (1993). 'Post Walrasian and Post Marxian economics', *Journal of Economic Perspectives*, 7 (1): 109–114.

Varian, H., (1984, 1992). *Microeconomic Analysis*, Norton, New York.

Veblen, T., (1898). 'Why is economics not an evolutionary science?', *Quarterly Journal of Economics*, 12: 373–397.

Veblen, T., (1899). 'The preconceptions of economic science', *Quarterly Journal of Economics*, 13: 121–150.

Veblen, T., (1909). 'The limitations of marginal utility', *Journal of Political Economy*, 17: 620–636.

Veblen, T., (1919 [1990]). W.J. Samuels (ed.), *The Place of Science in Modern Civilization*, Transaction Publishers, New Brunswick.

Whaples, R., (1995). 'Changes in attitudes among college economics students about the fairness of the market', *Journal of Economic Education*, 26: 308–313.

Wilber, C.K., (1996). 'Ethics and economics', in C.J. Whalen (ed.), *Political Economy for the 21st century*, M.E. Sharpe, Armonk.

Yao, X & Darwen, P., (2000). 'Genetic algorithms and evolutionary games', in W.A. Barnett, C. Chiarella, S. Keen, R. Marks, & H. Schnabl, (eds.), *Commerce, Complexity and Evolution*, Cambridge University Press, New York.

Yezer, A.M., Goldfarb, R.S. Poppen, P.J., (1996). 'Does studying economics discourage cooperation? Watch what we do, not what we say or how we play', *Journal of Economic Perspectives*, 10 (1): 177–186.

The price of everything and the value of nothing

Bishop, R.L., (1948). 'Cost discontinuities, declining costs and marginal analysis', *American Economic Review*, 38: 607–617.

Clapham, J.H., (1922a). 'Of empty economic boxes', *Economic Journal*, 32: 303–314.

Clapham, J.H., (1922b). 'The economic boxes–a rejoinder', *Economic Journal*, 32: 560–563.

Eiteman, W.J., (1947). 'Factors determining the location of the least cost point', *American Economic Review*, 37: 910–918.

Eiteman, W.J., (1948). 'The least cost point, capacity and marginal analysis: a rejoinder', *American Economic Review*, 38: 899–904.

Eiteman, W.J. & Guthrie, G.E., (1952). 'The shape of the average cost curve', *American Economic Review*, 42: 832–838.

Galbraith, J.K., (1997). *The Socially Concerned Today: the First*

Honorary Keith Davey Lecture, University of Toronto Press, Toronto.

Haines, W.W., (1948). 'Capacity production and the least cost point', *American Economic Review*, 38: 617–624.

Langlois, C., (1989). 'Markup pricing versus marginalism: a controversy revisited', *Journal of Post Keynesian Economics*, 12: 127–151.

Lee, F., (1996). 'Pricing and the business enterprise', in C.W. Whalen (ed.), *Political Economy for the 21st century*, ME Sharpe, Armonk.

Lee, F., (1998). *Post Keynesian Price Theory*, CambridgeUniversity Press, Cambridge.

Means, G.C., (1972). 'The administered-price thesis reconsidered', *American Economic Review*, 62: 292–306.

Pigou, A.C., (1922). 'Empty economic boxes – a reply', *Economic Journal*, 36: 459–465.

Pigou, A.C., (1927). 'The law of diminishing and increasing cost', *Economic Journal*, 41: 188–197.

Pigou, A.C., (1928). 'An analysis of supply', *Economic Journal*, 42: 238–257.

Robertson D.H., (1924). 'Those empty boxes', *Economic Journal*, 34: 16–31.

Robertson D.H., (1930). 'The trees of the forest', *Economic Journal*, 44: 80–89.

Robbins, L., (1928). 'The representative firm', *Economic Journal*, 42: 387–404.

Shove, G.F., (1930). 'The representative firm and increasing returns', *Economic Journal*, 44: 93–116.

Sraffa, P., (1926). 'The Law of returns under competitive conditions', *Economic Journal*, 40: 538–550.

Sraffa, P., (1930). 'The trees of the forest – a criticism', *Economic Journal*, 44: 89–92.

Shepherd, W.G., (1984). '"Contestability" vs. competition', *American Economic Review*, 74: 572–587.

Sato, K., (1979). 'Anote on capital and output aggregation in a general equilibrium model of production', *Econometrica*, 47: 1559–1568.

Size does matter

Keen, S., (2003). "Standing on the toes of pygmies: why econophysics must be careful of the economic foundations on which it builds", *Physica A*, 324: 108-116.

Keen, S., (2004). "Deregulator: Judgment Day for Microeconomics", *Utilities Policy*, 12: 109–125.

Keen, S., & Standish, R., (2009). "Profit Maximization, Industry Structure, and Competition: A critique of neoclassical theory", *Physica A* 370: 81-85.

Keen, S., & Standish, R., (2009). "Debunking the theory of the firm: a chronology",

American Review of Political Economy (www.arpejournal.com , forthcoming).

Martin, S., (2000). *Advanced Industrial Economics*, Basil Blackwell, Oxford.

Stigler, George J., (1957), "Perfect competition, historically contemplated". Journal of Political Economy, 65, pp. 1-17

Sato, K., (1979). 'Anote on capital and output aggregation in a general equilibrium model of production', *Econometrica*, 47: 1559–1568.

Shepherd, W.G., (1984). '"Contestability' vs. Competition'", *American Economic Review*, 74: 572–587.

The holy war over capital

Bhaduri, A, (1969). 'On the significance of recent controversies in capital theory: a Marxian view', *Economic Journal*, 79: 532–539.

Bharadwaj, K. & Schefold, B. (eds.), (1990). *Essays on Pierro Sraffa: Critical Perspectives on the Revival of Classical Theory*, Unwin Hyman, London.

Dumenil, G. & Levy, D., (1985). 'The classicals and the neo-classicals: a rejoinder to Frank Hahn', *Cambridge Journal of Economics*, 9: 327–345.

Hahn, F., (1982). 'The neo-Ricardians', *Cambridge Journal of Economics*, 6: 353–374.

Hodgson, G., (1991). *After Marx and Sraffa*, St Martin's Press, New York.

Samuelson, P., (1966). 'A Summing Up', *Quarterly Journal of Economics*, 80: 568–583.

Samuelson, Paul A., (1998). 'Report card on Sraffa at 100', *European Journal of the History of Economic Thought*, 5: 458–67.

Salvadori, N. & Steedman, I., (1988). 'No reswitching? No switching!',

Cambridge Journal of Economics, 12: 481–486.
Sraffa, P., (1960). *Production of Commodites by Means of Commodities: Prelude to a Critique of Political Economy*, CambridgeUniversity Press, Cambridge.

There is madness in their method

Albert, D.Z., (1992). *Quantum Mechanics and Experience*, HarvardUniversity Press, CambridgeMA.
Albert, D.Z., (1994). 'Bohm's alternative to quantum mechanics', *Scientific American*, 270, No. 5: 32–39.
Bell, D. & Kristol, I., (1981). *The Crisis in Economic Theory*, Basic Books, New York.
Bond, N.W. (2000). 'Psychology: a science of many faces', in N.W. Bond & K.M. McConkey (eds.), *An Introduction to Psychological Science*, McGraw-Hill, Sydney.
Caldwell, B. (ed.), (1984). *Appraisal and Criticism in Economics: a Book of Readings*, Allen & Unwin, London.
Einstein, A., (1916 [1961]). *Relativity: the Special and the General Theory*, Random House, New York.
Eiteman, W.J. & Guthrie, G.E., (1952). 'The shape of the average cost curve', *American Economic Review*, 42: 832–838.
Friedman, M., (1953). 'The methodology of positive economics'. Reprinted in B. Caldwell (ed.), (1984), *Appraisal and Criticism in Economics: a Book of Readings*, Allen & Unwin, London.
Hogan, C.J., Kirshner, R.P. & Suntzeff, N.B., (1999). 'Surveying space-time with supernovae', *Scientific American*, 280, No. 1: 28–33.
Kuhn, T., (1962). *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago.
Lakatos, I., (1978). *The Methodology of Scientific Research Programs*, J. Worrall & G. Currie (eds.), Cambridge University Press, New York.
Langlois, C., (1989). 'Markup pricing versus marginalism: a controversy revisited', *Journal of Post Keynesian Economics*, 12: 127–151.
Mirowski, P., (1984). 'Physics and the "marginalist revolution"', *Cambridge Journal of Economics*, 8: 361–379.
Musgrave, A., (1981). '"Unreal assumptions" in economic theory: the untwisted', *Kyklos*, 34: 377–387, reprinted in Caldwell (1984).

Let's do the Time Warp again

Blatt, J.M., (1983). *Dynamic Economic Systems*, ME Sharpe, Armonk.
Blaug, M., (1998). 'Disturbing currents in modern economics', *Challenge!*, 41(3): 11–34.
Braun, M., (1993). *Differential Equations and their Applications*, Springer-Verlag, New York.
Debreu, G., (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, YaleUniversity Press, New Haven.
Clark, J.B. 1898). 'The future of economic theory', *Quarterly Journal of Economics*, 13: 1–14).
, (1939). 'An Essay in Dynamic Theory', *Economic Journal*, 49: 14–33.
Hicks, J.R., (1949). 'Mr Harrod's Dynamic Theory', *Economica*, 16: 106–12.
Jorgenson, D.W., (1960). 'A dual stability theorem', *Econometrica*, 28: 892–899.
Jorgenson, D.W., (1961). 'Stability of a dynamic input-output system', *Review of Economic Studies*, 28: 105–116.
Jorgenson, D.W., (1963). 'Stability of a dynamic input-output system: a reply', *Review of Economic Studies*, 30: 148–149.
Keen, S., (1998). 'Answers (and questions) for Sraffians (and Kaleckians)', *Review of Political Economy*, 10: 73–87.
Keen, S., (2000). 'The nonlinear dynamics of debt-deflation', in W.A. Barnett, C. Chiarella, S. Keen, R. Marks, & H. Schnabl (eds.), *Commerce, Complexity and Evolution*, Cambridge University Press, New York.
Keynes, J.M., (1923). A Tract on Monetary Reform, in *The Collected Works of John Maynard Keynes*, IV, (London: Macmillan, 1971).
Kydland, F.E. & Prescott, E.C., (1990). 'Business cycles – real facts and a monetary myth', *Federal Reserve Bank of Minneapolis Quarterly Review*, 14: 3–18.
Mas-Colell, A., (1986). 'Notes on price and quantity tatonnement dynamics', in H.F. Sonnenshein (ed.), *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, Berlin.
McManus, M., (1963). 'Notes on Jorgenson's model', *Review of Economic Studies*, 30: 141–147.
Nikaido, H., (1996). *Prices, Cycles and Growth*, MIT Press, CambridgeMA.

O'Hara, M. (1995). *Market Microstructure Theory*, Blackwell, Cambridge.

Ott, E., (1993). *Chaos in Dynamical Systems*, CambridgeUniversity Press, New York.

Samuelson, Paul A. (1991). 'Apersonal view on crises and economic cycles' in M. Feldstein, (ed.), *The risk of Economic Crisis. A National Bureau of economic Research Conference Report*, Chicago and London: University of Chicago Press: 167–70.

Sargan, J.D., (1958). 'The instability of the Leontief dynamic model', *Econometrica*, 26: 381–392.

Steedman, I., (1992). 'Questions for Kaleckians', *Review of Political Economy*, 4: 125–151.

Taslim, F. & Chowdhury, A., *Macroeconomic Analysis for Australian Students*, Edward Elgar, Sydney.

Walras, L., (1874, 1900 [1954]). *Elements of Pure economics*, George Allen & Unwin, London.

The sum of the parts

Baird, W.C., (1981). *Elements of Macroeconomics*, West St Paul, New York.

Crouch, R., (1972). *Macroeconomics*, Harcourt Brace Jovanovitch, New York.

Dixon, R., (2000). *Walras Law and Macroeconomics*, www.ecom.unimelb.edu.au/ecowww/rdixon/wlaw.html.

Dixon, R., (2000). *A Formal Proof of Walras' Law*, www.ecom.unimelb.edu.au/ ecowww/rdixon/walproof.html.

Hicks, J.R., (1937). 'Mr Keynes and the "Classics": a suggested interpretation', *Econometrica*, 5: 147–159.

Hicks, J.R., (1980). 'IS-LM: an explanation', *Journal of Post Keynesian Economics*, 3 (2): 139–54.

Kates, S., (1998). *Say's Law and the Keynesian Revolution*, Edward Elgar, Cheltenham.

Keynes, J.M, (1925). 'A short view of Russia' in Keynes (1972), *Essays in Persuasion*, Cambridge University Press, Cambridge.

Keynes, J.M, (1936). *The General Theory of Employment, Interest and Money*, Macmillan, London.

Keynes, J.M, (1937). 'The general theory of employment', *Quarterly Journal of Economics*, 51: 209–223.

Klein, P.A., (1994). 'The reassessment of institutionalist mainstream relations', *Journal of Economic Issues*, 28: 197–207.

Leeson, R., (1991). 'The validity of the expectations-augmented Phillips curve model', *Economic Papers*, 10: 92–96.

Leeson, R., (1998). 'The origins of Keynesian discomfiture', *Journal of Post Keynesian Economics*, 20: 597–619.

Leijonhufvud, Axel, (1986). 'What would Keynes have thought of rational expectations?', in. J.L. Butkiewicz, K.J. Koford & J.B. Miller (eds.), *Keynes' Economic Legacy: Contemporary Economic Theories*, Praeger, New York: 25–52.

Leijonhufvud, Axel (1967 [1991]). 'Keynes and the Keynesians: a suggested interpretation' in E. Phelps (ed.), *Recent Developments in Macroeconomics*, Edward Elgar, Aldershot.

Marx, K., (1861 [1968]). *Theories Of Surplus Value, Part II*, Progress Press, Moscow.

Marx, K., (1885 [1956]). *Capital, Volume II*, Progress Publishers, Moscow.

Milgate, M., (1987). 'Keynes's General Theory', in J. Eatwell, M. Milgate & P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, Macmillan, London.

Minsky, H.M., (1963 [1982]). 'Can "it" happen again?', reprinted in H. Minsky, *Inflation, Recession and Economic Policy*, ME Sharpe, Armonk.

Minsky, H.M, (1975). *John Maynard Keynes*, ColumbiaUniversity Press, New York.

Say, J.B., (1821 [1967]). *Letters to Mr Malthus on several subjects of political economy and on the cause of the stagnation of commerce to which is added a Catechism of Political Economy or familiar conversations on the manner in which wealth is produced, distributed and consumed in society*, Augustus M'Kelly, New York.

The price is not right

Barber, W.J. (ed.), (1997). *The Works of Irving Fisher*, Pickering and Chatto, London.

Bentham, J., *In Defence of Usury*, http://socserv2.socsci.mcmaster.ca:80/~econ/ugcm/3ll3/bentham/usur

Feher, D.C., (1999). *Debt Deflation: The Birth of a Concept & Its Development Over Time*, unpublished Honours thesis, University of Western Sydney.

Fisher, I., (1929). 'Transcript of an address of Professor Irving Fisher', in W.J. Barber (1997), Volume 10.

Fisher, I., (1930). *The Theory of Interest*, Macmillan, New York, reprinted in W.J. Barber (1997), Volume 3.

Fisher, I., (1932). *Booms and Depressions: Some First Principles*, reprinted in W.J. Barber (1997), Volume 10.

Fisher, I., (1933). 'The debt-deflation theory of great depressions', *Econometrica*, 1: 337–355.

Glassman, J.K. & Hassett, K.A., (1999). *DOW 36,000 : The New Strategy for Profiting from the Coming Rise in the Stock Market*, Times Business, New York.

Institute of Chartered Financial Analysts (eds.), (1991). *The Founders of Modern Finance: Their Prize-winning Concepts and 1990 Nobel Lectures*, Institute of Chartered Financial Analysts, Charlottesville.

Henwood, D., (1997). Wall *Street*, Verso, New York.

Jones, (1989). *God and the Moneylenders*, Basil Blackwell, Oxford.

Keen, (1995). 'Finance and Economic breakdown: modelling Minsky's financial instability hypothesis', *Journal of Post Keynesian Economics*, 17: 607–635.

Mackay, C., (1841). *Extraordinary Popular Delusions and the Madness of Crowds*, Crown Trade Paperbacks, New York; http://www.litrix.com/madraven/madne001.htm.

Sharpe, W.F., (1964). 'Capital asset prices: a theory of market equilibrium under conditions of risk', *Journal of Finance*, 19: 425–442, reprinted in Institute of Chartered Financial Analysts (eds.) (1991).

Sharpe, W.F., (1970). *Portfolio Theory and Capital Markets*, McGraw-Hill, New York.

Strange, S., (1997). *Casino Capitalism*, ManchesterUniversity Press, Manchester.

Finance and economic breakdown

Barber, W.J. (ed.) (1997, *The Works of Irving Fisher*, Pickering and Chatto, London.

Barnett, W., (1999). 'Asingle-blind controlled competition among tests for nonlinearity and chaos', *Journal of Econometrics*, 82: 157–192.

Feher, D.C., (1999). *Debt Deflation: The Birth of a Concept & Its Development Over Time*, unpublished Honours thesis, University of Western Sydney.

Fisher, I., (1933). 'The debt-deflation theory of great depressions', *Econometrica*, 1: 337–355.

Glassman, J.K. & Hassett, K.A., (1999). *DOW 36,000 : The New Strategy for Profiting from the Coming Rise in the Stock Market*, Times Business, New York.

Institute of Chartered Financial Analysts (eds.), (1991). *The Founders of Modern Finance: Their Prize-winning Concepts and 1990 Nobel Lectures*, Institute of Chartered Financial Analysts, Charlottesville.

Haugen, R.A., (1999a). *The Beast on Wall Street, Prentice-Hall, New Jersey.*

Haugen, R.A., (1999b). *The New Finance*, Prentice-Hall, New Jersey.

Haugen, R.A., (1999c). *The Inefficient Stock Market*, Prentice-Hall, New Jersey.

Henwood, D., (1997). Wall Street, Verso, New York.

Kariya, T., (1993). *Quantitative Methods for Portfolio Analysis: MTV Model Approach*, Kluwer Academic Publishers, Dordrecht.

Keen, S., (1995). 'Finance and economic breakdown: modelling Minsky's financial instability hypothesis', *Journal of Post Keynesian Economics*, 17: 607–635.

Mackay, C., (1841). *Extraordinary Popular Delusions and the Madness of Crowds*, Crown Trade Paperbacks, New York; http://www.litrix.com/madraven/madne001.htm.

Minsky, H.M., (1970). 'Financial instability revisited: the economics of disaster', in Minsky (1982), 117–161.

Minsky, H.M., (1977). 'The financial instability hypothesis: an interpretation of Keynes and an alternative to 'standard' theory', *Nebraska Journal of Economics and Business*, reprinted in Minsky (1982), 59–70.

Minsky, H.M., (1980). 'Capitalist financial processes and the instability of capitalism', *Journal of Economic Issues*, 14, reprinted in Minsky (1982), 71–89.

Minsky, H.M., (1982). *Inflation, Recession and Economic Policy*, Wheatsheaf, Sussex.

Ott, E., (1993). *Chaos in Dynamical Systems*, CambridgeUniversity Press, New York.

Peters, E.E., (1994). *Fractal Market Analysis*, John Wiley & Sons, New York.

Peters, E.E., (1996). *Chaos and Order in the Capital Markets (2nd edition)*, John Wiley&Sons, New York.
Strange, S., (1997). *Casino Capitalism*, ManchesterUniversity Press, Manchester.

Don't shoot me, I'm only the piano

Barbier, E.B. (ed.), (1993). *Economics and Ecology*, Chapman & Hall, London.
Caplan, B., (2000). 'Rational irrationality: a framework for the neoclassical-behavioural debate', *Eastern Economic Journal*, 26: 191–212.
Costanza, R., (1993). 'Ecological economic systems analysis: order and chaos', in Barbier (1993), 29–45.
Franklin, J. & Daoud, A, (1988). *Introduction to Proofs in Mathematics*, Prentice-Hall, New York.

Nothing to lose but their minds

Böhm-Bawerk, (1896 [1949]). *Karl Marx and the Close of His System*, P. Sweezy (ed.), Orion Editions, New York.
Bose, A, (1980). *Marx on Exploitation and Inequality*, OxfordUniversity Press, Delhi.
Freeman, A & Carchedi, G., (1996). *Marx and Non-Equilibrium Economics*, Edward Elgar, Cheltenham.
Keen, S., (1993a). 'Use-value, exchange-value, and the demise of Marx's labour theory of value', *Journal of the History of Economic Thought*, 15: 107–121.
Keen, S., (1993b). 'The misinterpretation of Marx's theory of value', *Journal of the History of Economic Thought*, 15: 282–300.
Keen, S., (2001). 'Minsky's thesis: Keynesian or Marxian?', in R. Bellofiore & P. Ferri (eds.), *The Legacy of Hyman Minsky*, Edward Elgar, Aldershot.
McLellan, D., (1971). *Karl Marx: Early Texts*, Blackwell, Oxford.
Marx, K., (1847). *Wage Labour And Capital*, Progress Publishers, Moscow.
Marx, K., (1857). *Grundrisse*, Penguin, Middlesex.
Marx, K., (1859). *A Contribution to the Critique of Political Economy*, Progress Publishers, Moscow.
Marx, K., (1861). *Theories Of Surplus Value, Parts I, II And III*, Progress Press, Moscow.
Marx, K., (1865 [1951]). 'Wages, price and profit' in *Marx-Engels Selected Works, Volume I*, Marx-Engels-Lenin Institute (ed.), Foreign Languages Publishing House, Moscow.
Marx, K., (1865). *Wages, Price And Profit*, Progress Publishers, Moscow.
Marx, K., (1867, 1885, 1894). *Capital, I, II & III*, Progress Publishers, Moscow.
Marx, K., (1879 [1971]). 'Marginal notes on A. Wagner', in D. McLennan (ed.), *Karl Marx: Early Texts*, Basil Blackwell, Oxford.
Meek, R., (1972). *The Economics of Physiocracy*, George Allen & Unwin, London.
Mohun, S., (1994). *Debates in Value Theory*, St Martin's Press, New York.
Rosdolsky, R., (1977). *The Making of Marx's Capital*, Pluto Press, London.
Shaikh, A, (1982). 'Neo-Ricardian economics: a wealth of algebra, a poverty of theory', *Review of Radical Political Economics*, 14: 67–83.
Smith, A., (1776 [1838]). *An Inquiry into the Nature and Causes of the Wealth of Nations*, Adam and Charles Black, Edinburgh.
Sweezy, P.M., (1942). *The Theory of Capitalist Development*, OxfordUniversity Press, New York.
Wilde, L., (1989). *Marx and Contradiction*, Avebury, Aldershot.

There are alternatives

Arestis, P., Dunn, S.P. & Sawyer, M., (1999). 'Post Keynesian economics and its critics', *Journal of Post Keynesian Economics*, 21: 527–549.
Barnett, W.A, Chiarella, C., Keen, S., Marks, R. & Schnabl, H. (eds.), (2000). *Commerce, Complexity and Evolution*, CambridgeUniversity Press, New York.
Bowles, H. & Gintis, S., (1993). 'The revenge of Homo Economicus: contested exchange and the revival of political economy', *Journal of Economic Perspectives*, 7 (1): 83–102.
Caldwell, B.J. & Boehm, S. (eds.), (1992). *Austrian Economics: Tensions and New Directions*, Kluwer Academic, Boston.
Downward, P. & Reynolds, P., (1999). 'The contemporary relevance of Post-Keynesian economics: editors' introduction', *Economic*

Issues, 4: 1–6.

Harcourt, G., (1982). *The Social Science Imperialists*, (Kerr, P., ed.), Routledge & Kegan Paul, London.

Hodgson, G.M, (1998). *The Foundations of Evolutionary Economics*, Edward Elgar, Cheltenham.

Hodgson, G.M, (1999). *Evolution and institutions: On Evolutionary Economics and the Evolution of Economics*, Edward Elgar, Cheltenham.

Hodgson, G.M, Samuels, W.J. & Tool, M.R. (eds.), (1994). *The Elgar Companion to Institutional and Evolutionary Economics*, Edward Elgar, Aldershot.

Kirzner, I.M., (1996). *Essays on Capital and Interest*, Edward Elgar, Cheltenham.

Kregel, J.A., (1983). 'Post-Keynesian theory: an overview', *Journal of Economic Education*, 14 (4): 32–43.

Kydland, F.E. Prescott, E.C., (1990). 'Business cycles – real facts and a monetary myth', *Federal Reserve Bank of Minneapolis Quarterly Review*, 14.

Lee, F.S., (1998). *Post Keynesian Price Theory*, CambridgeUniversity Press, New York.

Levins, R. & Lewontin, R.C., (1986). *The Dialectical Biologist*, HarvardUniversity Press, London.

Littlefield, S. (ed.), (1990). *Austrian Economics*, Edward Elgar, Aldershot.

Palley, T.I., (1996). *Post Keynesian Economics: Debt, Distribution and the Macro Economy*, Macmillan, London.

Nightingale, J., (2000). 'Universal Darwinism and social research: the case of economics', in W.A. Barnett, C. Chiarella, S. Keen, R. Marks & H. Schnabl (eds.), *Commerce, Complexity and Evolution*, Cambridge University Press, New York.

Rassuli, A. & Rassuli, K.M, (1988). 'The realism of Post Keynesian economics: a marketing perspective', *Journal of Post Keynesian Economics*, 10: 455–473.

Rosser, J.B., (1999). 'On the complexities of complex economic dynamics', *Journal of Economic Perspectives*, 13 (4): 169–192.

Rothbard, M, (1963, 1972). *America's Great Depression*, Sheed&Ward, Kansas City.

Rothbard, M, (1970). Power and the Market, Sheed Andrews and McMeel Inc., Kansas City.

Rotheim, R.J., (1999). 'Post Keynesian economics and realist philosophy', *Journal of Post Keynesian Economics*, 22: 71–103.

Sawyer, M.C. (ed.), (1988). *Post-Keynesian Economics*, Edward Elgar, Aldershot.

Sciabarra, C., (1995). *Marx, Hayek and Utopia*, State University of New York Press, New York.

Tisdell, C., (1995). 'Evolutionary economics and research and development', in S. Dowrick (ed.), *Economic Approaches to Innovation*, Edward Elgar, Aldersot.

Walters, B. & Young, D., (1997). 'On the coherence of Post Keynesian economics', *Scottish Journal of Political Economy*, 44: 329–349.

Witt, U. (ed.), (1992). *Explaining Process and Change: Approaches to Evolutionary Economics*, University of Michigan Press, Ann Arbor MI.

---

[1] This is actually an application of the 'theory of the second best' (Lancaster & Lipsey 1956). Briefly, Lancaster and Lipsey showed that any single step towards what economics describes as the ideal situation could reduce welfare, if more than one step was required to move from the present situation to the ideal.

[2] *Finance and Economic Breakdown*, to be published by Edward Elgar Publishers. I will report progress towards this book on my website www.debtdeflation.com.

[3] This is not to suggest that Stiglitz is unaware of criticisms of what he labels 'smart economics' in this article. See Stiglitz 1993.

[4] It is also because complexity theory tends to be incompatible with neoclassical economics, since a common property of complex systems is that they have unstable equilibria: see Chapters 8 and 12.

[5] Nonlinear 'difference' equations also generate chaos, but economics courses normally cover only linear difference equations.

[6] Given the greenhouse effect, it is quite possible that some day meteorologists will argue that we must control the environment in order to

avoid calamities. But at present, meteorologists are more inclined to tell us to stop doing things that alter the climate.

[7] There are at least four other definable schools of thought (see Chapter 14), and it is in one sense an impressive achievement of neoclassical economics that it alone is identified as 'economics' in the public eye.

[8] If you're sceptical about this, the analogy I give in Chapter 8 explains why statics and dynamics can be in conflict.

[9] Except in one footnote, where the equation concerns meteorology rather than economics, and can easily be skipped. Occasionally, when some proposition in the text is best stated in mathematical form, I have use words rather than mathematical symbols.

[10] Smith's statement was actually made in the context of the advantages of domestic investment over foreign investment-though he also argued that unregulated trade was superior to regulated trade. To see Smith's statement in context, follow the links to Hedonism/invisible.

[11] Most of Bentham's endeavours in this regard related to devising a scale of punishments that he regarded as just sufficient to discourage the commission of crime.

[12] Jevons, one of the three co-founders of neoclassical economics, was justly sceptical that mathematics could treat all behaviour. He argued that "economy does not treat of all human motives. There are motives nearly always present with us, arising from conscience, compassion, or from some moral or religious source, which economy cannot and does not pretend to treat. These will remain to us as outstanding and disturbing forces; they must be treated, if at all, by other appropriate branches of knowledge" (Jevons 1866). However subsequent economists have applied this theory to all behaviour, including inter-personal relations.

[13] Cardinal refers to the ability to attach a precise quantity, whereas ordinal refers to the ability to rank things in size order, without necessarily being able to ascribe a numeric value to each.

[14] Economists assume that consumers spend all their income. They treat savings, in effect, as a form of consumption-only what is consumed is goods in the future.

[15] The 'curve' is actually a straight line here, but it doesn't have to be-it's just the easiest curve to draw, and the one economists themselves almost always draw.

[16] While the theory apparently derives the individual demand curve from the individual's indifference map, in fact the demand curve is more 'observable' than the indifference map.

[17] Or the "rate of time preference'.

[18] As your income rises, your expenditure on ground and instant coffee may change so that you cease buying instant coffee. Your tastes can also alter in a way in which economists would represent by drawing a new set of indifference curves.

[19] The result was in fact first proved by Gorman in 1953, but the proof was neglected until its rediscovery by Sonnenschein in the early 70s.

[20] Kreps 1990 is one honourable exception.

[21] However in a multi-commodity world, where demand for one product depends upon demand for another, perverse results are possible: a demand curve for a single commodity can therefore be upward sloping in price on occasions.

[22] Say gives a typical statement (reproduced on the web at Hedonism/Say) of this approach to utility, which denies the ability of anyone to judge or measure the utility any other individual garners from a particular commodity.

[23] Kirman's paper is an eloquent and well-argued instance of the phenomenon that those who have constructed the 'high theory' of economics are far less confident about its relevance than more ordinary economists.

[24] Smith and Ricardo allowed exceptions to this rule; neoclassical economics in effect made these exceptions the rule.

[25] Money is simply a measuring stick in this analysis, and the monetary unit could as easily be pigs as dollars.

[26] Marketing expenses cannot be added in to 'rescue' the doctrine, since the true purpose of marketing is to alter the firm's demand curve, and this only makes sense if firms produce differentiated products-something the

theory of perfect competition explicitly rules out.

[27] The 'cost curve' for any one firm or industry is the product of interactions between all industries, an issue that is ignored in the neoclassical treatment of a single market. This issue is discussed in Chapter 6.

[28] Economics ignores the issue of ecological sustainability, though clearly it must be considered by a reformed economics.

[29] Assuming for the sake of argument that we can provide meaningful formulas for the change in revenue and costs over time. In reality of course, the future is uncertain, which is yet another source of weakness in neoclassical economics.

[30] However, the proof should be obvious to mathematically aware readers.

[31] Though empirical work suggests that in practice, there is little sign of any negative relationship between the quantity sold and the price-and hence little evidence of a 'demand curve' (Lee 1996).

[32] The situation is even worse when we acknowledge that economic advice about how to maximise profits is categorically wrong in a dynamic, uncertain economy.

[33] An alternative model, known as 'contestability', argues that a monopoly can have the same properties as a competitive industry if it is subject to costless entry by new firms. I do not consider this theory here because the critiques of perfect competition in this chapter apply equally to contestability.

[34] Economists are likely to deflect these critiques by arguing that the theory has moved well beyond the simplistic models taught to undergraduates. However, at the very least economists should stop teaching these models. Secondly, economists still regard the model of perfect competition as describing the ideal economy. This chapter argues that this ideal is in fact a farce.

[35] This assumption is of course inconsistent with the assumption of a 'short run', during which some factor of production cannot be changed, which is essential to get the phenomenon of diminishing marginal productivity, that in turn generates a rising marginal cost. Firms already inside the industry are assumed to be unable to alter their capital equipment at all, but in the same time period firms not currently in the industry can build a factory and move in on the market? Hello? This logic is about as watertight as the script of the average daytime soap opera. But we can allow even this nonsensical assumption and still show that the theory of perfect competition is illogical.¬

[36] The material in this section is new, and has yet to be published in an economics journal (though it has been published in the physics journal *Physica A*; see Keen 2003 and Keen & Standish 2006). Similar but more complex arguments have been put by Sato (1979) (and also Barnett 1979), who basically concluded that aggregation is only possible if all firms are perfectly competitive, and that labor's share in output is the same across all industries. This is equivalent to assuming that there is only one industry – in which case aggregation is meaningless.

[37] The linear demand curve for the figures shown here had a price of 800 for zero output, and a price of zero for an output level of 800,000,000 units; while the marginal cost function for each firm was set so that the market marginal cost curve was given by a parabola with a marginal cost of 100 per unit for zero output, rising to 1200 for an industry-level output of 10^{10} units.

[38] Sraffa's 1926 paper critiqued a forerunner to this idea: that economies of scale could be external to the firm, but internal to the industry. This would mean that as an industry expanded, all firms benefited from lower costs, but none benefited any more than any other. Sraffa argued that few, if any, economies of scale would fit in this category. Instead most would be internal to a firm, and thus advantage the big firm more than the small.

[39] Economics has a theory of markets that are dominated by a few large firms, which is known as oligopoly theory. I have not discussed it in this book because it takes for granted many of the concepts that I have already debunked-such as diminishing marginal productivity and smooth, downward-sloping demand curves. However a critique of this area will appear in a future edition of this book.

[40] Though inequality in America fell somewhat in the later stages of its 'Bubble Economy', as unemployment dropped below 5%. While the redistributive effects of the post-Bubble period will be complex, it is highly unlikely that the poor will see their absolute positions improve.

[41] Only the downward sloping, positive segment is relevant, of course.

[42] If the x axis represents bananas, then the x-intercept shows how many

bananas can be bought with the consumer's entire income.

[43] Normal, that is, if we ignore the problems pointed out with the concept of the supply curve in chapters 3 and 4.

[44] The interpretations are superficially similar, until you consider a wage that is below the starvation level. The economic interpretation would show a worker working more hours still. A more realistic expectation is that workers would attempt theft or revolution.

[45] In the real world, a really high wage will encourage those who wish to accumulate money for its own sake to 'work their butts off'. Economists can't avail themselves of such an explanation, since they argue that the sole purpose of earning an income is to enable consumption.

[46] While this case is most easily made with equations, I'll stick to words here.

[47] The same case can be made with respect to the change in the wages bill, but I focus just on profit times capital to keep the argument simple.

[48] The ratio of a change in capital to a change in capital is 1.

[49] This will apply only when the capital to labour ratio is the same in all industries – which is effectively the same as saying there is only one industry.

[50] The rule in this example is that 10 quarters of wheat had to exchange for 1 ton of iron, or 2 pigs. These are relative prices-ratios in which commodities exchange – rather than absolute prices in terms of money.

[51] At least, not until a 'von Neumann machine'-a machine that can both produce output and reproduce itself-is invented.

[52] This is often all economists know of Sraffa's critique, and they dismiss it immediately by saying that it wrongly ignores the issue of marginal productivity. In fact, there is much more to Sraffa's critique, and Bhaduri's critique establishes the invalidity of the assertion that the rate of profit equals the marginal productivity of capital.

[53] When output is measured in terms of a 'standard commodity', and when the wage is normalised so that when the rate of profit r is zero, the wage w equals 1.

[54] This correspondence is not exact, but it can be made accurate to any level short of 100% by continuing the process of reduction for long enough.

[55] I'm enough of a wine buff to realise that this example is practically impossible, but it will do as an illustration.

[56] I first heard this joke in a public debate between my then Professor of Economics and a physicist. I now appreciate the irony that physicists are turning their attention to economics – and in general being horrified by neoclassical economic theory.

[57] I am grateful to my student Marchessa Dy for suggesting this very evocative analogy.

[58] The analysis below is a brief summary of Imre Lakatos's concept of competing 'scientific research programs'. The philosophy of science is today dominated by more 'post-modernist' concepts. I will leave exploration of these newer strands to the interested reader to pursue.

[59] Ironically, Austrian economics, an alternative schools of thought that is very closely related to neoclassical economics, differs by singing the praises of capitalism as a disequilibrium system (see Chapter 14).

[60] Equilibrium in turn has been endowed with essential welfare properties, with a 'Pareto
optimal equilibrium' being a situation in which no-one can be made any better off without making someone else worse off.

[61] If you have ever taught a child to ride a bike, you would know that this lesson is the most difficult one to grasp-that a moving bike balances itself, without the need for training wheels or other props which would keep it upright when it was stationary.

[62] This analogy is apt in more ways than one. The art of balancing a stationary bike requires great skill, and anyone who has mastered it is likely to 'show it off' at every opportunity?regardless of how impractical it might be. Similarly, economists who have mastered the difficult mental gymnastics involved in equilibrium analysis take every opportunity to parade their prowess-regardless of how irrelevant this skill might be to the art of managing a real economy.

[63]Only the gold market in London even approaches this structure, and even that is a market at which only one commodity is traded, rather than 'all commodities' (O'Hara 1995).

[64]See the explanation of difference and differential equations in Chapter 12.

[65]The main theorem is the Perron-Frobenius theorem on the eigenvalues of a positive matrix.

[66]Debreu used a notation that allowed for negative prices and negative quantities. However this was a convenience only, and has no impact on the analysis in this section.

[67]The actual equations were: 'the rate of change of x with respect to time equals the constant a multiplied by (y-z); the rate of change of y with respect to time equals x multiplied by (b-z) minus y; the rate of change of z with respect to time equals (x multiplied by y) minus (c multiplied by z)'.

[68]The two equations are linked, because workers' wage demands depend on the rate of employment, while investment-which determines the rate of growth-depends on income distribution (a higher workers' share means lower profits, and hence lower investment).

[69]These models are the subject of my next book, *Finance and Economic Breakdown*.

[70]Keynes lumped what we today term neoclassical economists with those we today call the classical economists. While they are distinctly different schools of thought, Keynes was correct to group them together on this issue, since they concurred that a general deficiency of aggregate demand was impossible.

[71]Surprisingly few books give this argument in full, given the extent to which it is a core belief in economics. Two that do are W. C. Baird, (1981), Elements of Macroeconomics, West St Paul (Chapter 3), and R. Crouch, (1972), Macroeconomics, Harcourt Brace Jovanovitch, New York (Chapter 6).

[72]His actual procedure was to argue that, when employment increased, demand for consumer goods would increase by less than the increase in employment, and that equilibrium would only be achieved if investment demand automatically took up the slack. This confusing argument is equivalent to the simpler case set out here.

[73]As Milgate observed, 'Received opinion, that Keynes's General Theory is a contribution to 'disequilibrium' analysis, was stamped indelibly upon the collective consciousness of the economics profession at an early date-by critics and converts alike' (Milgate 1998).

[74]Marx's theory of value is normally regarded as the labour theory of value, which is criticised in Chapter 13. I argue that his theory of value is something quite different.

[75]Hicks also had savings depending upon the level of output, but output was already determined by the first equation and therefore "we do not need to bother about inserting Income here unless we choose" (Hicks 1937).

[76]Total employment is the sum of the number of workers needed to produce investment output and the number needed to produce consumption output, so if labour productivities differ between the two sectors then the breakdown has to be known before total employment is determined.

[77]The basic idea is that a change in investment causes a more than proportional change in income. Aggregate investment is the 'independent variable'; consumption, on the other hand, is merely a passive residual. An increase in investment adds to income directly, and also indirectly by its effect on consumption-because workers have to be hired, and suppliers paid, to carry out the investment. They in turn spend part of what they are paid, so that an increase in investment causes a more than proportional increase in expenditure and hence income-the 'multiplier'.

[78]Defined as a fall of more than 10% from the most recent peak.

[79]Defined as a fall of more than 20% from the most recent peak.

[80]Though he avoided bankruptcy thanks to loans from his wealthy sister-in-law (they were never repaid, and she forgave them in her will; Barber 1997), and selling his house to Yale in return for life tenancy.

[81]Fisher's theory was first published in another work in 1907; *The Rate of Interest* restated this theory in a form which Fisher hoped would be more accessible than was the 1907 book.

[82] Muslim societies continue with the traditional definition, and therefore prohibit-with varying degrees of effectiveness-any loan contract in which the lender does not share in the risk of the project..

[83] In a country, such as Great Britain, where money is lent to government at three per cent and to private people upon a good security at four and four and a half, the present legal rate, five per cent, is perhaps as proper as any." (Smith, 1776, Book II, Chapter 4)

[84] The best record of the famous early panics is in Charles MacKay's *Extraordinary Popular Delusions And The Madness Of Crowds*. The chronicler of our day is Charles P. Kindleberger.

[85] Of course, many of the high-flying companies of 1929 were no longer in the index in 1955, so that anyone who held on to their 1929 share portfolio took far more than 25 years to get their money back, and most of the shares they held were worthless.

[86] Barber notes that after Fisher came into great wealth when his filing invention was taken over by the Remington Rand Corporation, he was "eager to add to his portfolio of common stocks and placed himself in some exposed positions in order to do so. At this time, his confidence in the soundness of the American economy was complete" (Barber 1997).

[87] Barber observed that among the other reasons was the fact that "In the 1930s, his insistence on the urgency of 'quick fix' solutions generated frictions between Fisher and other professional economists" (Barber 1997).

[88] Almost 90% of the over 1200 citations of Fisher in academic journals since 1956 were references to his pre-Great Depression works (Feher 1999).

[89] Since diversification reduces risk, all investments along this edge must be portfolios rather than individual shares. This concept is important in Sharpe's analysis of the valuation of a single investment, which I don't consider in this summary.

[90] I have slightly modified Sharpe's figure so that C's indifference curve is tangential to the IOC at F.

[91] In words, this formula asserts that the expected return on a share will equal the risk-free rate (P), plus 'beta' times the difference between the overall market return and the risk-free rate. Beta itself is a measure of the ratio of the variability of a given share's return to the variability of the market index, and the degree of correlation between the share's return and the market index return.

[92] There are three variations on this, known as the weak, semi-strong and strong forms of the EMH.

[93] Barnett's test used 3.57 (the 'Feigenbaum number') rather than 4, and his random series was scaled to avoid obvious anomalies like negative values.

[94] Chaos was first 'discovered' by Henri Poincar– in 1899, when he tried to find a solution to the 'many body problem'-the problem of gravitational attraction between a star and more than just one planet-and instead proved instead that there was no analytic solution; instead, the bodies would follow complex aperiodic paths (i.e., cycles occur which never exactly repeat themselves, unlike conventional cyclical functions like sine waves, etc.) which were later labelled 'chaotic'.

[95] More complex data distributions are predicted by some more elaborate versions of the EMH, but the normal distribution is still the overall yardstick.

[96] There are a number of econometric analyses that attempt to account for this. As Peters comments, they capture some of the local statistical features, but fail to capture the overall characteristics (Peters 1994).

[97] Haugen is a proponent of 'behavioural finance', which has been gaining acceptance in applied and academic finance in recent years, though its adherents are still a minority compared to supporters of the efficient markets hypothesis.

[98] Though early in the cycle debt to equity and interest to profit ratios may in fact come in lower than anticipated, due to the multiplier effect of generally increased investment, and the consequently larger profits from each investment.

[99] Ponzi's 'bank' grew rapidly (before its spectacular demise) with its high interest rates financed out of new deposits. During the Florida land boom that preceded the 1929 Wall Street crash, he sold as many as 23 lots to the acre, advertised as being near a city that did not exist. (Minsky, 1977 [1982]; 1978 [1982]; Galbraith 1988)

[100] The suggestion was made on the 'Post Keynesian Thought' internet discussion list, and in a paper delivered to the workshop on expectational and learning dynamics in financial markets at the University of Technology Sydney in December 1999. It was published in the Norwegian newspaper Aftenposten ('The Evening Post') on 14 April 2000.

[101] This proof, though mathematical, is very easy to understand. Follow the link to Maths/Irrational to see the proof.

[102] The most recent example I have seen of this is the paper by Caplan (2000) which attempts to explain findings – similar in spirit to those discussed in the Web page adjunct to Chapter 2 – which show that experimental subjects do not conform to the neoclassical definition or rational. Rather than accepting that the neoclassical definition of rationality may be flawed (again, for the reasons discussed on the web), Caplan proposes that irrationality may be a 'good', which people 'consume' like any other, and then represents a rationality-irrationality trade-off using indifference curves. This one article is not the final word on the neoclassical response to such findings. But I expect it to be far more readily adopted by the profession than any acknowledgement that, for example, the 'curse of dimensionality' makes rational behaviour as economists define it simply impossible in the real world.

[103] Though this branch of mathematics provides many tools which enables mathematicians to characterise the behaviour of more complex and realistic models of the real world – including such things as differential equation models of the economy.

[104] Though economists from several other schools of thought still pay great attention to Marx's original writings on economics, and see Marx as the father of many important concepts in economic analysis.

[105] This story may or may not be apocryphal. Check the website http://thebeadsite.com/FRO-MANG.html for one perspective, and http://www.crazyhorse.org/ for another.

[106] Sraffa's critique of the concept of an upward sloping demand curve, and the critiques of the market demand curve covered earlier, also undermine the neoclassical position and support the classical view.

[107] The subject was a bone of contention from the time of Aristotle on. However, predecessors to the physiocrats were quite unsystematic about the determination of value and price.

[108] This means that as output rose, costs of production fall. Smith was thus thinking in terms of a 'downward sloping supply curve'-at least in the medium to long term-in contrast to the upward sloping supply curve that is so central to economics today, which was debunked in chapter 3.

[109] All these examples were hypothetical, of course: Ricardo did not go out and measure the labour involved in producing the means of production in any industry, and then present his findings.

[110] "Not only his labouring servants, but his labouring cattle, are productive labourers" (Smith 1776).

[111] "By the invention of machinery ... a million of men may produce double, or treble the amount of riches, ... but they will on no account add anything to value" (Ricardo 1817). Marx commented that "This is quite wrong. The value of the product of a million men does not depend solely on their labour but also on the value of the capital with which they work." (Marx 1862).

[112] Marx qualified this as 'socially necessary labour-time', to take account of the possibility of out-of-equilibrium situations in which more labour-time might be lavished on a product than could be recouped by its sale.

[113] There is no reason why the rate of surplus value should be constant over time in practice, and Joan Robinson used this as the basis of her critique of Marxian economics. She argued that an increase in c could cause a rise in s/v, the rate of surplus value, so that the rate of profit would not fall over time.

[114] There were several counter-tendencies that could attenuate this, but ultimately Marx thought the tendency of the rate of profit to fall would prevail.

[115] This is an extremely brief outline of a much more complicated argument. Its purpose is not to provide a detailed exposition of Marx's theory of revolution, but to prepare the ground for critiques of the labour theory of value.

[116] This is clearly unrealistic, but the logic is the same even if we incorporate the reality that corn would be needed to produce corn. Steedman's example just made the numerical algebra easier to follow. He

then continued his argument using symbolic linear algebra, to establish the generality of his analysis.

[117] If I had worked with exact numbers rather than rounded them to two decimal places, the two calculations would have corresponded exactly. The value calculations, on the other hand, differ systematically, and by far more than can be attributed to rounding error.

[118] Similar arguments had been made before, as early as at the end of the 19th century. Steedman simply provided the most comprehensive and definitive critique.

[119] I dispute Bose's reading of Marx on this subject, but find the logic impeccable in his 'essence of value' analysis.

[120] He also employed a set of axioms from which his conclusions were derived.

[121] At each step in the reduction, one period's capital inputs are reduced to the previous period's direct labour and capital inputs, marked up by the equilibrium rate of profit.

[122] Services such as a massage, which might appear to be a commodity-free good, involve commodities directly (massage bench, oil), and if even these are foregone (an oil-free massage while lying on bare ground), they involve it indirectly through the need for the masseur to eat to stay alive. The commodity 'massage' could therefore not be reproduced in the absence of commodity inputs, such as food.

[123] Marx's philosophy was derived from Hegel's, with Marx arguing that he replaced Hegel's idealism with realism. Dialectics is popularly known as the trio of thesis-antithesis-synthesis, and though this concept is popularly associated with both Hegel and Marx, it in fact derives from another, lesser known philosopher Fichte. For an intelligent discussion of dialectical philosophy in general, and Marx's application of it in particular, see Wilde 1989.

[124] In a different type of economy use-value could well be brought to the foreground: commodities could be produced for the ruling elite at ostentatious expense, without regard to their cost of production. I well remember seeing a backscratcher in the ForbiddenPalace in Beijing, made out of jade, gold, diamonds, emeralds and rubies.

[125] This 'discovery' of the application of dialectical philosophy to economics occurred after Marx happened to re-read Hegel while he was drafting the Grundrisse (Oakley 1983, Mandel 1971).

[126] Marx's discussion of this example still attributed the increased surplus-value to labour, however the source of this difference was not any difference in the rate of surplus value with respect to labour employed, but to the postulate that the machine's use-value exceeded its exchange-value.

[127] "Exchange-value and use-value [are] intrinsically incommensurable magnitudes" (Marx 1867). Notice that Marx describes use-value as a magnitude in this circumstance. Outside production, when commodities are purchased to be consumed rather than being used to produce other commodities, their use-value will be qualitative, and therefore incommensurable with their exchange-values.

[128] This is a necessarily brief and personally opinionated survey of five very complex schools of thought. Readers who wish to delve deeper should consult the references given in this chapter. I have also omitted separate discussion of a notable school of economic thought, institutional economics, because I expect it to be subsumed under evolutionary economics.

[129] Though evolutionary theorists themselves now argue that Darwin's vision of the evolutionary process, in which 'nature did not make jumps', is flawed, and that therefore Darwinism is an inappropriate label for modern evolutionary theory (Schwartz 1999).

[130] And, ironically, some evolutionary theorists are now arguing that biological evolution may in some ways be purposive (McFadden 2000).

[131] There will be resistance aplenty too from government departments, and the bureaucracies of central banks, where promotion has come to those who have held the economic faith..